NASA-TM-82995

NASA Technical Memorandum 82995

Computer Program to Minimize Prediction Errors in Models From Experiments with 16 Hypercube Points and Zero to Six Center Points

Arthur G. Holms
Lewis Research Center
Cleveland, Ohio

November 1982

NASA

DISPLAY 97/2/1

83N21872*# ISSUE 11 PAGE 1779 CATEGORY 65 RPT#: NASA-TM-82995
E-1428 NAS 1.15:82995 82/11/00 64 PAGES UNCLASSIFIED DOCUMENT

UTTL: Computer program to minimize prediction error in models from experiments with 16 hypercube points and 0 to 6 center points

AUTH: A/HOLMS, A. G.

CORP: National Aeronautics and Space Administration. Lewis Research Center, Cleveland, Ohio. AVAIL.NTIS SAP: HC A04/MF A01

MAJS: /*COMPUTER PROGRAMS/*ERRORS/*FACTORIAL DESIGN/*OPTIMIZATION/*REGRESSION ANALYSIS

MINS: / PREDICTION ANALYSIS TECHNIQUES/ STATISTICAL DISTRIBUTIONS

ABA: Author

ABS: A previous report described a backward deletion procedure of model selection that was optimized for minimum prediction error and which used a multiparameter combination of the F - distribution and an order statistics distribution of Cochran's. A computer program is described that applies the previously optimized procedure to real data. The use of the program is illustrated by examples.

ENTER:

CONTENTS

N83-21822#

COMPUTER PROGRAM TO MINIMIZE PREDICTION ERRORS IN MODELS FROM EXPERIMENTS WITH 16 HYPERCUBE POINTS AND ZERO TO SIX CENTER POINTS

Arthur G. Holms

National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio 44135

SUMMARY

An important and efficient experiment is provided by the full or fractional two-level factorial experiment with 16 factorial points and zero to six center points. For such experiments, a previous report described a backward deletion procedure of model selection that was optimized for minimum prediction error and which used a multiparameter combination of the F-distribution and an order statistics distribution of Cochran's. A computer program is now described that applies the previously optimized procedure to real data. The use of the program is illustrated by examples.

Prior art was defined as using the F-distribution with backward deletion, testing terms in the nondecreasing order of their absolute magnitudes. Such a procedure is here called an F-only strategy. Typically, backward deletion has been used with a nominal significance level chosen to control the probability of Type 1 errors at a nominal level of 5 percent. The 5-percent level F-only strategy produced enormous increases in prediction errors over the recommended (multiparameter) strategies. A 50-percent level F-only strategy still produced small increases in prediction errors over the recommended strategies.

INTRODUCTION

Whereas experiments that elucidate the mechanism of a process should usually be preferred over experiments that merely give empirical data, experiments that result in empirical mathematical models often provide the quickest way to optimize a process, a material composition, or a structure. Techniques sometimes called Box-Wilson methods or, more recently, response

surface methods have been found to so systematize empirical investigations that an efficient approach to optimum conditions is assured.

In materials and structures development, the experiments are often characterized by

(1) Costly experimental units such as cast or forged specimens.
(2) Time-consuming techniques such as fatigue or rupture testing.
(3) Error-accumulating processes such as high-temperature melting or forging and testing in corrosive environments.

The two-level, fixed-effects, full- or fractional-factorial designs of experiment, without replication, are often appropriate for these situations. Some examples of work that fall within the preceding framework are described in Collins, Quigg, and Dreshfield (1968); Sandrock and Holms (1969); Kent (1972); and Filippi (1974).

Subset regression procedures are often used to select the terms that are retained in a multivariable polynomial equation that has been fitted to the results of an experiment. The selection or rejection of terms is usually done according to a statistical test at some arbitrary level of significance. The usual purpose of the final model is to predict the value of the dependent variable as a function of values of the independent variables that may be of interest. For such a purpose, the selection of terms should be done to minimize prediction error rather than to achieve an arbitrary level of significance.

The purpose of this report is to describe and illustrate the use of a previously optimized method (Holms (1980)) for the selection of multivariable polynomial models to be fitted to the results of two-level factorial experiments having zero to six center points. (The method was optimized to minimize the maximum prediction error over the points of the experiment.) A computer program that uses the method and its application to real experiments is illustrated by two examples. The context of the development of the method and its justification are outlined in appendix A.

SEQUENTIAL DELETION

Population Model

The single observation value of the response is assumed to occur according to the model

$$y = E(Y) + e \tag{1}$$

where the error e is independently normally distributed with mean zero and variance σ^2. (Some robustness against nonnormality for a chain-pooling procedure was demonstrated by Holms and Berrettoni (1967).)

For relatively saturated experiments smaller than 16 observations, the opinion is offered that such experiments are too small to provide both good estimates of model coefficients and a good test statistic in cases where random errors are large enough to call for a statistical decision procedure. The simulations of Holms (1980) were all performed with experiments containing 16 hypercube points plus zero to six center points in the belief that such experiments are large enough to justify the use of a statistical decision procedure but small enough so that the precise optimization of the procedure would be quite beneficial. Where g is the number of independent variables and the experiment is assumed to be a 2^{-h} fractional replicate of the full-factorial experiment, factorial observations are assumed to result one-for-one from the hypercube points and their number is

$$n_c = 2^{g-h} = 16$$

An example of an appropriate model equation for the population mean value of the response in the case of four independent variables is

$$\begin{aligned} E(Y) = {} & \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_1 x_2 + \beta_5 x_3 + \beta_6 x_1 x_3 + \beta_7 x_2 x_3 \\ & + \beta_8 x_1 x_2 x_3 + \beta_9 x_4 + \beta_{10} x_1 x_4 + \beta_{11} x_2 x_4 + \beta_{12} x_1 x_2 x_4 \\ & + \beta_{13} x_3 x_4 + \beta_{14} x_1 x_3 x_4 + \beta_{15} x_2 x_3 x_4 + \beta_{16} x_1 x_2 x_3 x_4 \end{aligned} \tag{2}$$

Model Fitting

The model is assumed linear with orthogonality of parameter estimates provided by the design of the experiment. An equation such as (2) is assumed fitted to the results of a two-level experiment where the x's are design values, namely, the high level of x_k is represented by $x_k = +1$, and the low level of x_k is represented by $x_k = -1$. (If center points are used, they have coordinates with all $x_k = 0$.) The initial model fitting is assumed to give least-squares estimates of the model parameters that are minimum variance unbiassed linear estimates, and for parameters beyond β_1, the estimates have the form

$$b_i = \frac{1}{n_c} \sum_{k=1}^{n_c} a_{ik} y_k \tag{3}$$

where $i = 2, \ldots, n_c$ and the a_{ik} are appropriate values of ± 1. Such estimates have population mean values

$$E(b_i) = \beta_i \tag{4}$$

Combination Estimate for Zero Degree Coefficient

A weighted estimate of the β_1 of equation (2) is to be formed from the n_c hypercube observations and the n_0 center point observations where all observations are assumed to have variance σ^2. Model coefficients estimated from the hypercube observations each have variance

$$V(b_i) = \sigma^2/n_c$$

Thus, the variance of the function estimate for a model such as equation (2) with coefficients all estimated, for example, by Yates' method, from such observations is (at the design center)

$$V(\hat{Y}_0) = V(b_1) = \sigma^2/n_c$$

Let Y_0 be estimated from a combination of the n_c hypercube points and the n_0 center points. Let $\bar{y}_0$ be the arithmetic mean of the n_0 center point observations. Then the estimate of Y_0 weighted inversely as the variances of b_1 and $\bar{y}_0$ is

$$\hat{Y}_0 = (n_c b_1 + n_0 \bar{y}_0)/(n_c + n_0)$$

Because the coefficient estimates are uncorrelated, the weighted estimate $\hat{Y}_0$ is also the least-squares estimate of β_1. Thus, if b_1 is the estimate of the zero degree coefficient from the Yates analysis, the least-squares estimate from the combined observations is

$$b_1^* = (n_c b_1 + n_0 \bar{y}_0)/(n_c + n_0) \tag{5}$$

Mean Squares and Sums of Squares

The squares of the estimates multiplied by n_c provide the numerator mean squares used in the hypothesis testing:

$$Z_i = n_c b_i^2 \tag{6}$$

These mean squares have expectations

$$E(Z_i) = \sigma^2 + n_c \beta_i^2 \tag{7}$$

where

$$V(Y_k) = \sigma^2 \tag{8}$$

for $i = 2, \ldots, n_c$ and $k = 1, \ldots, n_c$. Thus, from equation (7), if any β_i is zero, the associated Z_i is an estimator of σ^2.

Hypothesis Testing

The denominator for the hypothesis testing is based on the construction of sums of squares. Six cases are identified according to combinations of the values of n_0 and m_p, where n_0 is determined by the design of the experiment and m_p is chosen according to a strategy of hypothesis testing.

The cases are identified by the first three columns of table 1(a). Equations for the initial sum of squares to be used in the starting denominator of the test statistics are derived in appendix B and given in the fifth column of table 1(a).

The following remarks apply to the cases listed in table 1(a).

Because case a provides no denominator sum of squares, there can be no deletion procedure. All the model coefficients are estimated and all the terms are retained.

Case d uses $n_0 = 0$ and $m_p > 0$. This is the case investigated by Holms and Berrettoni (1969). The deletion method of that investigation was as follows.

The mean squares Z_i from the usual Yates analysis (aside from the zero degree coefficient) are ordered in nondecreasing magnitude and renamed $Z_{(j)}$:

$$Z_{(1)} \leq \cdots \leq Z_{(n)}$$

where $j = 1, \ldots, n$ and $n = n_c - 1$. Chain-pooling assumes that m_p of the smallest $Z_{(j)}$ have been generated with zero population coefficients. Their sum is called R_{j-1} where initially

$$j - 1 = m_p$$

Multiplication of the critical points of Cochran's (1941) distribution by j gives the critical points of the U_j-distribution tabulated by Holms and Berrettoni (1967). The mean square $Z_{(j)}$ is tested for significance at nominal level α using the statistic

$$U_j = jZ_{(j)} \Big/ \left[R_{j-1} + Z_{(j)}\right] \tag{9}$$

If $Z_{(j)}$ is not significant, j is indexed upward by one and the next mean square is tested. If the k^{th}-ordered mean square is the smallest mean square to test significant at level α, then all terms associated with smaller mean squares are deleted from the model.

Consider cases b and c where $n_0 > 0$ and $m_p = 0$. The first mean square to be tested is $Z_{(1)}$, and the null hypothesis is

$$H_0: \ \beta_{(1)} = 0 \tag{10}$$

where, for any j, $\beta_{(j)}$ is the parameter associated with the ordered mean square $Z_{(j)}$. The alternative hypothesis is

$$H_a: \ \beta_{(1)} > 0$$

Because U_j is not defined for $j < 2$, $Z_{(1)}$ cannot be tested against the U_j-distribution. If the test is performed against the F-distribution, the fact that $Z_{(1)}$ is an ordered statistic implies that a test of nominal size α will not have true size α. With this proviso, a nominal size α-test is performed.

For case b, the initial test statistic is

$$F_{1,ndfb} = \frac{n_{dfb} Z_{(1)}}{SS_b} \tag{11}$$

with n_{dfb} defined by equation (B3) and SS_b computed by equation (B4).

For case c, the equivalent initial test statistic is

$$F_{1,ndfc} = \frac{n_{dfc} Z_{(1)}}{SS_c} \tag{12}$$

where n_{dfc} is given by equation (B6) and SS_c is given by equation (B5). If $Z_{(1)}$ is reported significant, no further testing is done and there is no conditional deletion of terms.

For either case b or c, let SS_0 be the initial sum of squares and n_{df0} the initial degrees of freedom. If $Z_{(1)}$ was reported as insignificant, then it is pooled with SS_0 for a test of $Z_{(2)}$. The test statistic is then

$$F_{1,ndf0+1} = \frac{(n_{df0} + 1) Z_{(2)}}{SS_0 + Z_{(1)}} \tag{13}$$

Testing and pooling continue in this manner provided that insignificance is the result of the prior test.

The test statistic for any $Z_{(j+1)}$ is thus

$$F_{1,ndf0+j} = \frac{(n_{df0} + j)Z_{(j+1)}}{SS_0 + Z_{(1)} + \dots + Z_{(j)}} \tag{14}$$

For $j > 1$, the option exists of testing $Z_{(j)}$ against the F-distribution or against the U_j-distribution. These options are also both available for the first test of a $Z_{(j)}$ in cases d, e, and f; however, testing against the F-distribution might not be good for case d, because in case d there is neither a pure error nor a residual sum of squares, and the testing is performed entirely with ordered mean squares.

Suppose the situation is that of $n_0 > 0$ and $j > 1$. A criterion is needed for choosing between testing against the F-distribution or the U_j-distribution. If j is relatively small and n_0 is relatively large, the F-distribution might be more appropriate; whereas, if j is relatively large and n_0 is relatively small, the U_j-distribution might be more appropriate. The approach was for $j > 1$ to compute j/n_0 and use the F-distribution for $j/n_0 \leqq r_F$ and use the U_j-distribution for $j/n_0 > r_F$, where $0 \leqq r_F < \infty$ and r_F was optimized from Monte Carlo studies. Table 1(b) shows how the choice of the value of r_F affects the values of n_0 and j at which a transfer occurs from the use of the F-distribution to the use of the U_j-distribution.

Consider the case

$$j/n_0 > r_F \tag{15}$$

The U_j statistic is defined by equation (9). Suppose the criterion $j > r_F n_0$ has been met and the information in $y_{01}, \dots, y_{0n0}$ is to be combined with that in $Z_{(1)}, \dots, Z_{(j)}$ for a test of $Z_{(j)}$ against a critical point of the U_j-distribution. An approximation to equation (9) is

$$t_j = \frac{(n_{df0} + j)z_{(j)}}{SS_0 + z_{(1)} + \dots + z_{(j)}} \tag{16}$$

The distribution of t_j of equation (16) is merely an approximation of the distribution of U_j of equation (9) because the denominator in equation (16) has been stabilized by the n_{df0} mean squares in SS_0.

Under the null hypothesis, $z_{(j)}$ is the largest of an ordered sample of j estimators of σ^2. Although the quantity $(SS_0 + z_{(1)} + \dots + z_{(j)})/(n_{df0} + j)$ is an estimator of σ^2 as is the quantity $(r_{j-1} + z_{(j)})/j$ of

equation (9), the quantity j of equation (16) rather than the quantity $(n_{df0} + j)$ was used as the entry point of the U_j-tables because the tables are based in part on the numerator $z_{(j)}$ being the j^{th} extreme value of a sample of mean squares together having a mean value of σ^2.

<u>For case d</u>, with $n_0 = 0$, no testing or deletion can be done unless $m_p > 0$. The initial value of j is $m_p + 1$. Thus, even if $\alpha_F < 1.0$, with $m_p > 0$, j is > 1, and with $n_0 = 0$, j is $> r_F n_0$, and testing against the F-distribution is excluded. Thus, any testing with $n_0 = 0$ is done against the U_j-distribution, which implies that the first test using equation (9) takes place with

$$r_{j-1} = SS_d$$

where SS_d is defined by equation (B7) and $j = m_p + 1$. Subsequent testing is done with

$$t_j = \frac{jz_{(j)}}{SS_d + z_{(m_p+1)} + \ldots + z_{(j)}}$$

<u>For either case e or f</u>, with $j \leqq r_F n_0$, the test statistic is provided by equation (14) with SS_0 and n_{df0} being given by equations (B9) and (B10) for case e and by (B11) and (B12) for case f.

Terms Deleted

If $Z_{(j)}$ tests as insignificant, it is pooled with (added to) the denominator of the test statistic and j is indexed upward one unit. When a $Z_{(j)}$ tests as significant, testing is stopped and $\hat{\eta}$ terms are deleted from the model where $\hat{\eta}$ is the integer value of $r_\eta(j - 1)$ and r_η has been empirically optimized. If $Z_{(j)}$ tests as insignificant at $j = n_c - 1$, then $\hat{\eta}$ is the integer value of $r_\eta(n_c - 1)$. The terms deleted are those corresponding to the $\hat{\eta}$ smallest mean squares.

Definition of Strategy

In summary, the expressions for degrees of freedom and sums of squares are given in table 1(a), and the entire sequential deletion strategy is specified by the parameters (m_p, r_F, α_F, α_U, r_n). Where $0 \leqq m_p < n_c$, m_p is the number of mean squares initially pooled. If $Z_{(j)}$ is the j^{th} ordered mean square being tested, then $Z_{(j)}$ is tested against the U_j-distribution at nominal level α_U if $j > 1$ and $j > r_F n_0$. The mean square $Z_{(j)}$ is otherwise tested against the F-distribution at nominal level α_F. The convention $\alpha_F = 1.0$ is used to signify that no testing is done against the F-distribution, and the convention of $\alpha_U = 1.0$ is used to signify that no testing is done against the U_j-distribution. The number of terms found to be insignificant is multiplied by r_n and the integer value of the product (namely, $\hat{n}$) is the number of smallest absolute value coefficients whose terms will be deleted from the model. (The coefficient b_1^* of the zero degree term is excluded from the deletion procedure.) In the notation for the strategy parameter set (m_p, r_F, α_F, α_U, r_n), a long dash is used to represent a parameter if it has been made inoperative by a value assigned to some other parameter.

Selection of Strategy Parameters

Some of the more useful strategies investigated by Holms (1980) are summarized in table 1(c). The type a strategy should be used when the error is known to be small (coefficient of variation in the vicinity of 4.0 percent). The type b strategy should be used when the relative magnitude of the error is unknown. The type c strategy should be used where the relative error is known to be large (coefficient of variation as large as 65 percent).

COMPARISON WITH MORE USUAL METHODS

How do these multiple parameter strategies compare with the usual strategy? An attempt to answer the question requires some assumptions as to what constitutes a usual strategy. Commonly, a subset regression procedure

is used with coefficient deletion at an arbitrarily defined level of significance. As developed in appendix A, very little work has been done to identify procedures that minimize prediction error. The following assumptions defining a usual strategy are now introduced:

(1) The deletion procedure is based exclusively on the critical points of the F-distribution.

(2) The procedure uses stepwise sequential deletion, testing terms in the nondecreasing order of their absolute magnitudes.

(3) There is no arbitrary reintroduction of insignificant terms as implied by values of $r_\eta < 1.0$.

(4) There is no arbritrary initial deletion of terms as implied by values of $m_p > 0$. This assumption requires $\eta_0 > 0$.

The preceding assumptions imply that a usual strategy (henceforth, "F-only" strategy) can be compared with the proposed strategies by defining the F-only strategy as a strategy where (m_p, r_F, α_F, α_U, r_η) have the values (0, 16.0, α_F, 1.00, 1.0) and where α_F is chosen according to some criterion satisfactory to the analyst.

Results for the F-only strategies are given in table 1(d) and are compared with the multiparameter strategies in figure 1. The curves (fig. 1) for the F-only strategies always lie above the curves for the multiparameter strategies; that is, the F-only strategies are everywhere dominated by the multiparameter strategies and the F-only strategies are not admissible.

In addition to the coefficient of error functions $C_{ae,mx}(\nu)$ shown in table 1(d), regret functions $R(\nu)$ (defined in appendix A) are also listed. For values of η_0 from 1 to 6, the value of α that jointly minimizes both R(64.3) and R(4.0) is here called a security F-only or optimized F-only strategy. Table 1(d) shows that for all values of η_0 from 1 to 6 the optimizing value of α is 0.50. This result suggests that, whereas $\alpha = 0.05$ is a very popular value for significance testing, a value of 0.50 might often be preferred if the object is to minimize prediction error.

The values of $C_{ae,mx}(\nu)$ for the F-only strategy with $\alpha = 0.05$ and for the F-only strategy with $\alpha = 0.50$ are compared with the values of $C_{ae,mx}(\nu)$ for the security regret multiparameter strategy in table 1(e). For the more usual F-only strategy where $\alpha = 0.05$, the increase in prediction error over the range of η_0 was in the range of 19.2 to 28.9 percent at $\nu = 64.3$ and in the range of 157.3 to 901.8 percent at $\nu = 4.0$. When the F-only strategy

was used with the optimum α (namely, 0.50), the increase in prediction error varied from 4.7 to 8.8 percent at $\nu = 64.3$ (large coefficient of variation), and the change in prediction error varied from a decrease of 1.5 percent to an increase of 1.1 percent from the multiparameter values at $\nu = 4.0$ (small coefficient of variation).

USE OF COMPUTER PROGRAM

Application of the computer program (POOL10) is illustrated herein by two examples. A detailed description of the program input and output is given following the description of the second example. A description of the operating sections is given in appendix C, and the program listing is given in appendix D.

Illustrative Example: One-Half Replicate of 2^5 Experiment Without Center Points

The alloy development experiment that produced the data was described by Sandrock and Holms (1969). The levels of composition and pour temperature data are given in table 2(a) in which the levels of the independent variables are stated in natural units. The compositions melted are given in design units in table 2(b), and the mechanical property data are given in natural units.

One casting was made for each of the 16 compositions listed in table 2(b). Two stress-rupture tests from each composition were run at 1010° C and 103 MN/m^2 (1850° F and 15 ksi). The results are also shown in table 2(b). Testing variability and variability within castings could be observed from the test bar duplicates, but since there were no melt-to-melt duplicates, melt-to-melt variability could not be observed. A later investigation of melt-to-melt variability showed that the melt-to-melt variability significantly exceeded the bar-to-bar variability.

The next step is to subject the data to the model-fitting and deletion procedure of the computer program POOL10. The initial step is to use the logarithmic transformation to transform the values of the response t (stress-rupture life in hours) to logarithms of stress-rupture life y in order to generate a variable whose variance (scatter) would be approximately

constant over large changes in stress-rupture life that might result from the changes in composition.

The program reads out the input data and the mean value of the paired logarithms as illustrated by the values of Y(1), Y(2), and Y(I), respectively, in the illustrative output in appendix E. The model coefficients were then fitted by Yates' method, and sequential deletion procedures were invoked using the strategy parameters for $n_0 = 0$ listed in table 1(c). (The Yates method of the computer program attaches algebraic signs to the estimated coefficients as if the experiment were a full-factorial experiment on four independent variables. For the half-replicate experiment that was actually run, the signs that must be reversed are shown by the negative signs attached to the coefficients in table 7 of Holms (1967).)

Results are listed in table 3. The security regret strategy suggests that all terms of the model be retained; however, the three largest absolute value coefficients are all coefficients of first-degree terms. The strategy for large error suggests that two interaction terms $x_{Ti}x_{Cr}$ and $x_{Ti}x_T$ be retained; however, they have coefficients that are clearly smaller in absolute value than the coefficients of x_{Cr}, x_{Al}, and x_C. Furthermore, the absolute values of the coefficients of x_T and x_{Ti} are not much smaller than the absolute value of the coefficient of $x_{Ti}x_T$. Thus, the first-degree equation of the method of steepest ascents is a reasonable approximation to the data; namely,

$$\log t = 1.652 - 0.383x_{Cr} - 0.146x_{Al} + 0.100x_C + 0.037x_T - 0.030x_{Ti}$$

Illustrative Example: Full Replicate of 2^4 Experiment With Center Points

As described by Sandrock and Holms (1969), a vector experiment was designed and performed based on the results shown by the preceding equation. The results in the preceding equation and the results of the vector experiment were then used to design a second factorial experiment. It was performed without center point replication, and the conditions and results are shown in table 4.

The purpose of this section is to illustrate the use of the computer program POOL10 when center point data are available; however, no center point data were acquired until a star block (table V of Sandrock and Holms (1969)) had been performed. To use those center point data in this discussion, a mean value of the stress-rupture times $\bar{t}_C$ was computed for the observations of the hypercube experiment, and a mean value $\bar{t}_0$ was computed for the center point observations. The difference was used as an estimate of the block difference b_{sc} between the star block and the hypercube block as follows:

$$
\begin{aligned}
b_{sc} &= \bar{t}_0 - \bar{t}_c \\
&= 228.4 - 144.3 \\
&= 84.1
\end{aligned}
$$

The value 84.1 was then subtracted from the actual center point observations to obtain the adjusted center point observations used in the present example. The data from table V of Sandrock and Holms (1969) so adjusted are as follows:

Alloy	Stress-rupture life, hr
C-1	195.2, 185.6
C-2	114.3, 88.0
C-3	149.5, 119.8
C-4	158.7, 143.2

Without the use of center points, the coefficient estimates using the computer program POOL10 are as follows:

Term	Coefficient estimate	Term	Coefficient estimate
x^0	2.139	x_{Al}	-0.063
x_{Ti}	.002	$x_{Ti}x_{Al}$	-.054
x_{Cr}	.015	$x_{Cr}x_{Al}$	-.021
$x_{Ti}x_{Cr}$	-.004	$x_{Ti}x_{Cr}x_{Al}$	.025
x_C	.031	x_Cx_{Al}	.077
$x_{Ti}x_C$	.025	$x_{Ti}x_Cx_{Al}$	.031
$x_{Cr}x_C$	-.011	$x_{Cr}x_Cx_{Al}$	-.025
$x_{Ti}x_{Cr}x_C$	-.026	$x_{Ti}x_{Cr}x_Cx_{Al}$	.013

Because the coefficient of x^0 uses center point data in its estimate, its value changes with n_0. The values are as follows:

n_0	Coefficient of x^0
0	2.139
1	2.148
2	2.140
3	2.139
4	2.141

Because the two-level full- or fractional-factorial experiment provides orthogonal coefficient estimates, the estimates of coefficients that remain are not affected by the deletion of coefficients. The number of coefficients beyond the coefficient of x^0 is given as $\hat{\rho}$ by the deletion strategy, which means that the $\hat{\rho}$ largest absolute value coefficients and associated terms are to be retained in the model. The values of $\hat{\rho}$ for the different strategies and for differing possible numbers of center points are given in table 4(c).

As listed in table 4(c), the value of $\hat{\rho}$ for the security regret strategy for the largest value of n_0 was $\hat{\rho} = 6$. In the absence of any other information, this value of $\hat{\rho}$ should be accepted as giving the model equation with minimum prediction error. Using the coefficient of x^0 for $n_0 = 4$ with the other six largest absolute value model coefficients results in a predictive model as follows:

$$
\begin{aligned}
\log_{10} t = 2.141 &+ 0.031x_C - 0.026x_{Ti}x_{Cr}x_C \\
&- 0.063x_{Al} - 0.054x_{Ti}x_{Al} + 0.077x_C x_{Al} \\
&+ 0.031x_{Ti}x_C x_{Al}
\end{aligned}
$$

The form of the preceding equation shows very clearly that a first-degree equation cannot be fitted to the data and that the use of a vector experiment in the method of steepest ascents would be inappropriate. The next step should be the performance of a star block so that a quadratic model equation can be obtained for use with the method of local exploration, as was done by Sandrock and Holms (1969).

Description of Computer Program Input and Output

POOL10 is a FORTRAN-4 program that uses real data as input and performs the backward deletion using the statistical procedures developed in Holms (1980). The program is listed in appendix D, illustrative input is shown in figure 2, and illustrative output is given in appendix E. The following description of input and output has some paragraphs labeled identically with the applicable program sections in appendix D.

Section 1A - Declarations and tables. - The values of the nominal test size α are stored as (ALPHA(I), I = 1,11) and later used as output labels. These values range from 0.001 to 1.0; however, the value of 1.0 obtained by setting the index to 11 is merely a code implying that no significance testing is performed.

The sequential deletion requires critical values against which the test statistics are compared. The critical values of F are stored internally as ((FTB(I,J), J = 1,10), I = 1,20) where I indexes on the degrees of freedom and J indexes on the value assigned to α. The critical values of U_j are stored internally as ((TB(I,J), J = 1,10), I = 1,16) where I is the order number in nondecreasing order and J indexes on the value assigned to α.

Section 1B - Inputs and constants. - A logarithmic transformation of the dependent variates is provided. If KODE $\geqq$ 1, the logarithmic transformations are performed at the statement preceding statement 102.

Two-level, full or fractional experiments are sometimes performed in a manner that provides superficial rather than true replication. For example, if the treatments are each applied to different specimens but the dependent variable is observed through duplicate measurements on each of the specimens, such replication does not measure the specimen-to-specimen variability. A conservative procedure would be to compute the means of the repeated measurements and then proceed with these mean values as if they were observations from an experiment with no replication. The program computes the mean values in such cases of superficial replication for NYIN variates in each data set corresponding to a unique treatment. The operation is completed at statement 103.

The constants defining the input data and the sequential deletion strategy are read from data cards in the following order, with the order of

the fields being the same as the order of the symbols in the following description:

Format		Description
(13A6,A2)	REMARK (I)	Arbitrary literal information such as particular use of program and date of last change.
(2I1)	NO, KODE	NO is number of true center point replications. If KODE < 1, dependent variates are not transformed. If KODE $\geq$ 1, logarithmic transformation is used.
(A6,	DATAID	Treatment identification (may be omitted).
I6,	NYIN	Amount of superficial replication. Arithmetic mean of this number of variates is computed after any transformation and before model fitting.
6F6.0)	YIN	Response variates.
(3I5,2F5.3)	MP, KPF, KPU, RF, RETA	(The associated READ statement is actually in section 3.)

Illustrative input data are shown in figure 2. If the data are punched into data cards, the input is as follows. The first card gives the literal information in (13A6,A2) format stating that the computer run uses data from a particular experiment (namely that of table 4) and that the data were keypunched on 04-02-82. The second card in (2I1) format states that NO, the number of center point observations, is 4 and that KODE = 1, which implements the logarithmic transformation of the dependent variates. The next 16 cards in (A6, I6, 6F6.0) format give the basic information of the hypercube points. The first six columns in A6 format were used to reproduce the alloy designation of table 4(b). The next six columns in I6 format were used to record the number of superficial replicates NYIN, which in all cases was 2. Columns 13 through 48 are available in 6F6.0 format and in the present instance columns 13 through 18 were used to record the observation of 126.7 hours for alloy 1 of table 4(b), and columns 19 through 24 were used to record the observation of 176.5 hours for alloy 1 of table 4(b). Of course,

all 16 cards for the alloys numbered 1 through 16 are hypercube points and must be in the proper order for the operation of Yates' method of coefficient estimation. Following the information for the 16 hypercube observations are the cards containing the center point information in the same (A6, I6, 6F6.0) format as was used for the hypercube data. The number of such cards must be the same as was specified previously by NO.

Following the center point data cards are the strategy specifications in (3I5, 2F5.3) format where the value of MP is given in the first five columns, the value of KPF is given in columns 6 to 10, the value of KPU is given in columns 11 to 15, the value of RF (namely, r_F) is given in columns 16 to 20, and the value of RETA (namely, r_η) is given in columns 21 to 25. The relation between KPU or KPF and α_U or α_F is as follows:

KPU or KPF	α_U or α_F
1	0.001
2	.002
3	.005
4	.01
5	.025
6	.05
7	.10
8	.25
9	.50
10	.75
11	1.00

Thus, the last three cards of figure 2 specify strategies as follows:

m_p	r_F	α_F	α_U	r_η
0	0.5	0.75	0.50	0.85
0	.5	.50	.50	.80
0	.9	.10	.05	.80

Section 5 - Output. - On completion of the deletion of terms, the final step is writing the input and output data. Such data are illustrated in appendix E for the input data of figure 2. The first line of output merely repeats the literal information on the first data card. The next line gives the values of n_c, n_0, and KODE by means of the statement

NC = 16 NO = 4 KODE = 1

(KODE = 1 shows that a logarithmic transformation is used). Following that statement are column headings for the input data, namely, the literal information identifying the observation, followed by the observations, where there might be as many as six superficial replicates as indicated by the column headings Y(1), ..., Y(6). The input hypercube and center point observations are then listed in E 13.5 format under those column headings. Following that listing is the two-column array headed DATAID and Y(I), and under these column headings are the original literal data identifications and the mean value of the transformed observations (under Y(I)).

The statements NDF = 3 and S = 0.66830-01 give the values of n_{df} and s_1 in I3 and E 13.5 formats as computed prior to any sequential deletion. The next line of output gives the values of the strategy parameters (m_p, r_f, α_F, α_U, r_η) which, for the present example, are given by the line

MP = 0 RF = 0.500 ALPHAF = 0.500 ALPHAU = 0.500 RETA = 0.800

Values of $\hat{\rho}$, $\hat{\eta}$, and the smallest mean square concluded to be significant are given by the line

IRHO = 6 JETA = 9 Z(JETA + 1) = 0.11194-01

The column headings DATAID and B(I) indicate that the columns contain the original data identifications together with the coefficient estimates (however the coefficients might be aliased) in the order they are computed by Yates' method. These estimates reflect the consequences of the deletion procedure in that deleted coefficients have been set equal to .00000.

The last line gives the values of n_{df} and s_2 as estimated subsequent to the deletion procedure by the line

NDF = 15 S = 0.23759-01

The last three lines of figure 2 show three sets of strategy parameters (m_p, r_F, α_F, α_U, r_η) as input data. For every such input card there is a separate output line with the symbols

MP = RF = ALPHAF = ALPHAU = RETA =

and all the subsequent output as just described.

CONCLUDING REMARKS

A computer program was presented that combines the familiar Yates method of model coefficient estimation with a sequential backward deletion procedure. In contrast with many subset regression procedures, the deletion procedure does not attempt to reject terms according to some arbitrary significance level but, instead, is based on model-deletion strategies that were optimized for minimum prediction error. The optimization was done for two level factorial experiments containing 16 factorial points and any number from zero to six center points. As such, the procedure was optimized for situations containing very meager pure replication, and provision was made for the conditional use of insignificant coefficient estimates to improve the stability of the test statistic.

An attempt was made to define what might be called a prior art of sequential deletion for comparison with the recommended strategies. The prior art was assumed to use the F-distribution with backward deletion, testing terms in the nondecreasing order of their absolute magnitudes. Such a procedure is here called an F-only strategy. Typically, backward deletion has been used with a significance level chosen to control the probability of Type 1 errors at a nominal level of 5 percent. The present investigation showed that a good F-only strategy for minimizing prediction errors would use a nominal significance level of 50 percent.

The 5-percent level F-only strategy produced enormous increases in prediction errors over the recommended (multiparameter) strategies. The 50-percent level F-only strategy still produced small increases in prediction errors compared with the multiparameter strategies. In particular:

1. The 5-percent level F-only strategy increased prediction errors 20 to 30 percent at a large coefficient of variation and 150 to 900 percent at a small coefficient of variation over the prediction errors for the multiparameter strategies.

2. The 50-percent level F-only strategy (optimal among F-only strategies but seldom used) still increased prediction errors 5 to 9 percent at a large coefficient of variation but resulted in essentially no change at a small coefficient of variation over the prediction errors for the multiparameter strategies.

APPENDIX A

DEVELOPMENT OF DELETION PROCEDURE

Background

For two-level, fixed-effects factorial experiments without replication a suitable practice consists, according to Davies (1956), of pooling some arbitrary number of the highest order interaction mean squares into an estimate of error variance. When this is done, any of these high-order interactions that are not actually small or any unknown block effects (major changes in experimental conditions not accounted for by the model) could inflate some of the pooled interactions and, thereby, give too large an estimate of error variance. Too large an error estimate reduces the sensitivity of subsequent tests.

The preservation of sensitivity, when pooling mean squares into the estimate of error variance, has been an object of the procedure of Daniel (1959) and of Wilk, Gnanadesikan, and Freeny (1963).

Daniel (1959) uses the absolute values of the effect estimates as order statistics. They are plotted on probability paper, and the result is called a half-normal plot. In addition to conditional structuring of the model, Daniel's objectives included the determination of "bad values, heteroscedasticity, dependence of variance on mean, and some types of defective randomization, ..." The half-normal plot combined with a background of experience might provide a method by which a skillful user could pass judgment on the results of an experiment. Daniel concluded that such a plot can be used to make judgments about the reality of the largest effects observed only if a small proportion of the effect estimates represent real effects. Birnbaum (1959) investigated procedures related to half-normal plotting. His results on "...the probabilities of the various possible sorts of errors..." are limited to the single largest order statistic. He surmised that if only a small number of effect estimates have nonzero means, the power and sensitivity properties will tend to hold approximately.

The procedure of Wilk, Gnanadesikan, and Freeny (1963) if used with 2^ℓ treatments is benefited if some subjective or prior knowledge is used to decide that n of the $2^\ell - 1$ mean squares do not contain real effects or

that $\rho = 2^{\ell} - \eta - 1$ mean squares do contain real effects. As was shown by Wilk, Gnanadesikan, and Freeny (1963), their procedure is not robust against errors in guessing the value of η, and η must be guessed if prior knowledge is lacking. Chain pooling does not require prior knowledge of η. The procedure gives an estimate of η.

Zahn (1975) presented results of a Monte Carlo study of Daniel's original version of the half-normal plot and two modifications of it when as many as six real contrasts of size 1σ to 8σ are present in a 2^{p-q}, where $p - q = 4$, factorial experiment. Zahn considered the power, false positive behavior, and variance estimation of these versions of the half-normal plot when applied to the general problem in the case $n = 15$, because there are 15 contrasts of interest, ignoring the grand mean, in a 2^4 factorial experiment. He then pointed out that these versions or obvious modifications of them can be used for any n.

Zahn, Birnbaum, and Daniel limited their results to experiments where only a small proportion (at most 6 out of 15) of the effects are anticipated to be significant. On the other hand, situations can exist where the experimenter might use a two-level, fractional-factorial experiment designed such that most of the effects are significant. This situation is met by the procedures of this report.

If a two-level, full- or fractional-factorial experiment is performed and n_c observations are obtained from n_c orthogonal experimental conditions, the appropriate empirical equation for representing the results can have as many as n_c terms, each with a coefficient that has been fitted to the data. When this is done a question that should be asked is: Can the predictive accuracy be improved if some of the terms are deleted The fact that some of the terms might degrade the predictive accuracy of a fitted equation was recognized by Walls and Weeks (1969) who gave no procedure for identifying such terms.

A method for the sequential deletion of terms that was intended to reduce the prediction error was given by Kennedy and Bancroft (1971). Their method assumed that the experimenter has a prior established order for subjecting the terms to a sequence of significance tests. Unfortunately, in many experimental situations there is no subject matter basis for establishing a prior order, and in such cases an order-statistics procedure is appropriate. An order-statistics approach for significance testing was used

in the related papers by Daniel (1959) and Birnbaum (1959). They were not then seeking to minimize prediction errors.

For model-selection procedures used with small saturated experiments (experiments designed to have only as many experimental conditions as there are model parameters to be fitted), the analysis should begin with a minimum number of estimable terms being sacrificed to form a denominator for the test statistic. A procedure using m-terms sacrificed, where m can be as small as one, was investigated by Holms and Berrettoni (1969). The object was to delete terms in a manner where some control was maintained over the probabilities of Type 1 or Type 2 decision errors.

The minimizing of prediction error was the object of an investigation of a chain-pooling strategy as described by Holms (1974). Whereas that investigation had assumed only one cycle of analysis would be used, a suggestion by Holms and Berrettoni (1969) was that more than one cycle should be used. The purpose of a further chain-pooling investigation (Holms (1977)) was to optimize a combined procedure that might contain more than one analysis cycle where the procedure is to be optimized for minimum prediction error. An important application of chain pooling occurs in empirical optimum seeking.

A widely accepted methodology for the design and analysis of experiments efficient for the empirical attainment of optimum conditions was introduced by Box and Wilson (1951) and refined by Box and Hunter (1957). That method is now known as response surface methodology. Designs that are optimal for fitting second-degree equations were studied by Lucas (1974 and 1976), who was concerned with the optimality of single-block designs, but multiblock designs are often appropriate in the applications of response surface methods. A catalog of multiblock designs, limited to those particularly applicable to response surface methods, was given by Holms (1967). Response surface methodology assumes that hypercube and star blocks contain center points. Criteria for the numbers of such points to use, together with tables of recommended numbers, were given by Box and Hunter (1957). Criteria leading to much smaller numbers of center points for single-block experiments were given by Lucas (1976). The purpose of Holms (1979) was to characterize the experiment designer's options for numbers of center points in a range from very small to moderately large for multiblock sequential designs. The multiblock sequential designs were those for which treatment tables had been

given by Holms (1967). The numbers of center points used in each of the hypercube blocks ranged from zero to six.

As stated by Birnbaum (1959), the optimal decision procedure when $\rho \leq 1$ uses a test developed by Cochran (1941). The Z_i (originally in Yates' order) are ordered (omitting the mean square for the grand mean) according to nondecreasing magnitude and renamed $Z_{(j)}$:

$$Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(j)} \leq \cdots \leq Z_{(n)}$$

Cochran's statistic is

$$C_n = \frac{Z_{(n)}}{Z_{(1)} + \cdots + Z_{(n)}}$$

and the null hypothesis is rejected with test size α if C_n exceeds the upper 100α-percent point of Cochran's distribution.

As pointed out by Zahn (1975), the operating characteristics of the decision procedures that have been intended for $\rho \leq 1$ deteriorate rapidly with $\rho > 1$. But in addition to dealing with the situation of $\rho > 1$, the model-selection procedure must also deal with the possibility that not all the nonzero effects are of equal magnitude. As shown in table 8 of Zahn (1975), the detection rate for two means of size 4σ, 6σ, or 8σ is usually reduced in the presence of two other means of size 2σ, 4σ, or 6σ, respectively, from the value it would have been in the presence of two other means of the same size, namely, 4σ, 6σ, or 8σ, respectively. Because inequality of the means is clearly a disadvantage, the question arises as to what is the least favorable distribution of these means. As developed by Holms and Berrettoni (1969), a normal distribution of these means would be highly unfavorable.

Accordingly, the procedures of Holms (1980) were developed empirically to be optimal when the means do have an approximate normal distribution. Furthermore, the prior assumption of a normal distribution for these means in the general experimental situation does not seem to be unreasonable. Thus, the Monte Carlo founded procedures can be regarded as being empirically optimized against both what is highly likely to occur and what would be highly unfavorable if it did occur. Therefore, these procedures should be

regarded as being approximately both Bayes and minimax optimal. The procedures were developed to minimize the prediction errors in models fitted to the results of experiments having 16 hypercube points and any of zero to six center points. Results were exhibited as tables giving some of the operating characteristics of admissible strategies for each of the center point options. A security regret strategy was identified within each set of admissible strategies, and it was shown to be useful for a wide range of coefficients of variation.

Admissible Strategies

In the simulation studies of Holms (1980), the prediction errors e_{pi} of the deleted model equation were observed at each point of the hypercube, and the mean of the squared values $\bar{e}^2_{pi}$ was computed over the 1000 simulated experiments. The maximum of such error squares over the treatments was recorded as

$$\bar{e}^2_{max} = \max\, (\bar{e}^2_{pi}) \qquad i = 1, \ldots, n_c$$

The square root of $\bar{e}^2_{max}$ was adjusted for the scale factor θ of the simulations and for the number of treatments n_t where

$$n_t = n_c + n_0$$

using the equation

$$C_{ae,mx} = n_t^{1/2}\, \bar{e}_{max} / \theta$$

The quantity $C_{ae,mx}$ was then defined to be the adjusted coefficient of error. Its values were reported as functions of the scale parameter θ and as functions of the deletion strategy $(m_p, r_F, \alpha_F, \alpha_U, r_n)$.

The ratio (in percent) of the single observation standard deviation for the simulations to the population value of the function for the largest value of the dependent variable was defined as the coefficient of variation. Scale parameter values of θ of 0.125 and 2.0 resulted in coefficient of variation values of 64.3 and 4.0, respectively. In terms of laboratory

applications of response surface methods, these coefficients of variation were regarded as being large and small, respectively.

Many strategies were investigated to determine a strategy that minimized $C_{ae,mx}$ for each of the values of $n_0 = 0, 1, 2, 3, 4, 5, 6$ and for each coefficient of variation. Many strategies giving intermediate results were also investigated. A strategy was classed either as admissible or dominated according to its values of $C_{ae,mx}$ at both $\nu = 64.3$ and $\nu = 4.0$. A strategy was said to be dominated if for $\nu = 64.3$ there were another strategy with the same or lesser $C_{ae,mx}$ for $\nu = 64.3$ and a lesser $C_{ae,mx}$ at $\nu = 4.0$. A strategy was also said to be dominated if there were a strategy with the same or lesser $C_{ae,mx}$ at $\nu = 4.0$ and a lesser $C_{ae,mx}$ at $\nu = 64.3$. Any strategy that was not dominated was defined as being admissible. The strategies that were found to be admissible, together with some of their operating characteristics, were tabulated by Holms (1980).

Security Regret Strategy

The order of the tabulations for each value of n_0 in Holms (1980) was in the order of decreasing values of $C_{ae,mx}$ at $\nu = 64.3$, which is in the order of increasing values of $C_{ae,mx}$ at $\nu = 4.0$.

The regret function of a statistical decision procedure as a function of a parameter ν is here defined as the excess loss occurring with the procedure at a particular value of ν, as compared with the loss that would have occurred had the best statistical decision procedure been used for that particular value of ν. For the purposes of the present investigation, a regret function $R(\nu)$ is defined for $\nu = 64.3$ as being the $C_{ae,mx}(64.3)$ for any strategy divided by the value of $C_{ae,mx}$ for the best strategy for that value of ν; $R(\nu)$ is defined for $\nu = 4.0$ as being the $C_{ae,mx}(4.0)$ for any strategy divided by the value of $C_{ae,mx}$ for the best strategy for that value of ν.

Thus, for the successive values of n_0, the regret functions $R(n_0,\nu)$ are

$$R(n_0,\ 64.3) = C_{ae,mx}(n_0,\ 64.3)/\min[C_{ae,mx}(n_0,\ 64.3)]$$

and

$$R(n_0,\ 4.0) = C_{ae,mx}(n_0,\ 4.0)/\min[C_{ae,mx}(n_0,\ 4.0)]$$

From table 4 of Holms (1980), the values of $\min[C_{ae,mx}(n_0,\ 64.3)]$ and $\min[C_{ae,mx}(n_0,\ 4.0)]$ were as follows:

n_0	$\min[C_{ae,mx}(n_0,\ 64.3)]$	$\min[C_{ae,mx}(n_0,\ 4.0)]$
0	29.46	2.065
1	30.39	2.118
2	31.11	2.187
3	31.95	2.244
4	32.63	2.300
5	33.51	2.352
6	34.25	2.401

The single strategy that has the smallest regret function over both $\nu = 64.3$ and $\nu = 4.0$ is defined as the security regret strategy. The security regret strategy is thus the sequential deletion procedure which, for a given n_0, produces the least increase in prediction error (for $\rho = 15$ and an unfavorable distribution of parameters) over that prediction error which could have been achieved if the best strategy had been chosen for the given (unknown) value of error variance σ^2.

In examining the $R(\nu)$ values of table 4 of Holms (1980), for a given value of n_0, the parameters that give the security regret strategies are those that give the joint minimums on $R(64.3)$ and $R(4.0)$; these joint minimums were identified by asterisks. Thus, for the given values of n_0, the security regret strategies and the associated values of $C_{ae,mx}(\nu)$ are as follows:

n_0	m_p	r_F	α_F	α_U	r_n	$C_{ae,mx}(64.3)$	$C_{ae,mx}(4.0)$
0	1	---	1.0	0.50	0.25	32.95	2.240
1	0	3.0	.50	.10	.80	31.78	2.180
2	0	3.0	.25	.50	.85	33.14	2.252
3	0	.0	.75	.50	.80	33.29	2.301
4	0	.5	.50	.50	.80	33.89	2.309
5	0	1.0	.25	.10	.80	33.97	2.394
6	0	.5	.50	.05	.80	34.72	2.408

The question can be asked as to what choice of n_0 will result in the most efficient experiment. If the object of a choice of n_0 is to use the

most efficient choice together with a security regret strategy for deleting terms, then the preceding table shows that the most efficient choice (the choice that minimizes each of $C_{ae,mx}(64.3)$ and $C_{ae,mx}(4.0)$ is the choice of $n_0 = 1$). This choice applies to the condition of $n_c = 16$.

Selection of a Strategy

Results for the strategy that minimizes $C_{ae,mx}(4.0)$, for the security regret strategy, and for the strategy that minimizes $C_{ae,mx}(64.3)$ are presented in that order for each value of n_0 in table 1(c).

In summary, if the experimenter wishes to minimize the maximum prediction error over the 16 hypercube points of an experiment with n_0 center points when the variance error is relatively large (coefficient of variation in the range of 65 percent), the strategy for a given n_0 should be the last-listed strategy (for the given n_0) of table 1(c). If the experimenter wishes to minimize the maximum prediction error over the points of the experiment when the variance error is relatively small (coefficient of variation in the range of 4 percent), the strategy for a given n_0 should be the first-listed strategy (for the given n_0) of table 1(c).

If the experimenter has no basis for a choice of one of the two preceding extreme choices, the choice should be a security regret strategy as indicated by the b results in table 1(c) where (for all of the n_0 values) the largest value of the regret function was $R(64.3) = 1.1185$ as listed in table 1(c). This value of the regret function shows that for the worst value of $n_0 (n_0 = 0)$ the relative prediction standard error was increased by at most about 12 percent over what it would have been if the worst value of ν had occurred and the best strategy against it had been used. Thus, the security regret strategies (for each of the values of n_0) must be concluded to be widely useful strategies.

Estimates of Standard Deviation

Two procedures are used to estimate the single observation standard deviation. The first provides an estimate s_1 obtained prior to any model deletion. For the six cases listed in table 1(a) the estimates are computed according to

Case a: $s_1 = 0.0$
Case b: $s_1 = (SS_b)^{1/2}$
Case c: $s_1 = (SS_c/(n_0-1))^{1/2}$
Case d: $s_1 = (SS_d/m_p)^{1/2}$
Case e: $s_1 = (SS_e/(1+m_p))^{1/2}$
Case f: $s_1 = (SS_f/(n_0-1+m_p))^{1/2}$

As illustrated in appendix E, the number of degrees of freedom and the value of s_1 are printed following the printing of Y(I).

Subsequent to the termination of the sequential deletion procedure, the accumulated number of degrees of freedom n_{df} and the accumulated sum of squares SS are used to estimate a standard deviation as follows:

$$s_2 = (SS/n_{df})^{1/2}$$

The values of n_{df} and s_2 are printed following the listing of the coefficients B(I).

APPENDIX B

DERIVATION OF EQUATIONS FOR SUMS OF SQUARES

Case a ($n_0 = 0$, $m_p = 0$). - This case provides no information for a sum of squares for a test statistic.

Case b ($n_0 = 1$, $m_p = 0$). - Let Y_{0k} be the value observed at the k^{th} center point (origin) observation. Then, by the definition of σ^2,

$$V(Y_{0k}) = \sigma^2 \tag{B1}$$

Also from the definition of σ^2 where Y_i is the i^{th} hypercube observation,

$$V(Y_i) = \sigma^2 \tag{B2}$$

The object is to estimate σ^2 from the information in the y_i, $i = 1, \ldots, n_c$, and a single center point observation y_{01}. Because the model coefficient estimates in the two-level fractional-factorial experiment are orthogonal, the least-squares estimates of the regression coefficients from the combined data are all the same as the Yates estimates, except for the coefficient of the zero degree term b_1. Its least-squares estimate is from equation (5):

$$b_1^* = \left(y_{01} + \sum_{i=1}^{n_c} y_i\right) \Bigg/ (1 + n_c)$$

For any of the treatment points, let Δ_i be the difference between the observed value and the predicted value of Y where Δ_0 is the center point difference and $i = 0, 1, \ldots, n_c$. The Yates estimate of β_1 is

$$b_1 = \frac{1}{n_c} \sum_{i=1}^{n_c} y_i$$

The predicted values under the least-squares estimation are therefore all augmented by $b_1^* - b_1$ over their Yates method predictions.

The differences between the Yates method predictions and the observations are all zero at the hypercube points; therefore, over the $n_c + 1$ treatment points

$$\Delta_0 = y_{01} - b_1^*$$

$$\Delta_i = b_1 - b_1^* \qquad i = 1, \ldots, n_c$$

The estimate of σ^2 from the residual of the least-squares regression is

$$s^2 = \frac{\sum_{i=0}^{n_c} \Delta_i^2}{n_c + 1 - n_c} = (y_{01} - b_1^*)^2 + n_c(b_1 - b_1^*)^2$$

where

$$y_{01} - b_1^* = y_{01} - \frac{y_{01} + \sum_{i=1}^{n_c} y_i}{1 + n_c} = \frac{n_c y_{01} - \sum_{i=1}^{n_c} y_i}{1 + n_c}$$

and

$$b_1 - b_1^* = \frac{\sum_{i=1}^{n_c} y_i}{n_c} - \frac{y_{01} + \sum_{i=1}^{n_c} y_i}{1 + n_c} = \frac{\sum_{i=1}^{n_c} y_i - n_c y_{01}}{n_c(1 + n_c)}$$

Thus,

$$s^2 = (y_{01} - b_1^*)^2 + n_c(b_1 - b_1^*)^2 = \frac{n_c}{1 + n_c}(y_{01} - b_1)^2$$

For $n_c = 16$,

$$s^2 = 0.941176(y_{01} - b_1)^2$$

For this case, the number of degrees of freedom is

$$n_{dfb} = 1 \tag{B3}$$

and the sum of squares is

$$SS_b = s_1^2 n_{dfb} = 0.941176(y_{01} - b_1)^2 \tag{B4}$$

In this case, the error sum of squares was obtained from a residual involving b_1^*. This usage of b_1^* has the disadvantage that it will introduce a bias or lack-of-fit component into the sum of squares if the fitted model is biassed at the center point. Because of this bias risk, a pure error sum of squares will be computed if $n_0 > 1$.

Case c ($n_0 > 1$, $m_p = 0$). - Let the center point observations be y_{0k}; $k = 1, \ldots, n_0$. Their sample mean is

$$\bar{y}_0 = \frac{1}{n_0} \sum_{k=1}^{n_0} y_{0k}$$

and the sum of squares SS_c for case c is now:

$$SS_c = \sum_{k=1}^{n_0} (y_{0k} - \bar{y}_0)^2 = \sum_{k=1}^{n_0} y_{0k}^2 - \frac{1}{n_0}\left(\sum_{k=1}^{n_0} y_{0k}\right)^2 \tag{B5}$$

where the number of degrees of freedom n_{dfc} is

$$n_{dfc} = n_0 - 1 \tag{B6}$$

Case d ($n_0 = 0$, $m_p > 0$). - Let

$$SS_d = \sum_{j=1}^{m_p} z_{(j)} \tag{B7}$$

The number of degrees of freedom n_{dfd} is

$$n_{dfd} = m_p \tag{B8}$$

Case e ($n_0 = 1$, $m_p > 0$). - This is the additive situation of cases b and d:

$$SS_e = SS_b + SS_d \tag{B9}$$

$$n_{dfe} = n_{dfb} + n_{dfd} = 1 + m_p \tag{B10}$$

Case f ($n_0 > 1$, $m_p > 0$). - This case is additive with respect to cases c and d:

$$SS_f = SS_c + SS_d \tag{B11}$$

$$n_{dff} = n_{dfc} + n_{dfd} = n_0 - 1 + m_p \tag{B12}$$

APPENDIX C

OPERATING SECTIONS OF PROGRAM POOL10

The input (section 1) and output (section 5) of the program were described in the main text of this report. This appendix describes the operating sections of the program, namely, sections 2, 3, and 4, with sections 3 and 4 illustrated by the flow chart of figure 3. The program was tested on a Univac 1100 system.

Section 2 - Yates' method. - The algorithm for Yates' method is described as follows: The observations $y_{i,j}$ may be visualized as a column ($j = 1$) with row index $i = 1, \ldots, 2^{\ell}$. The column is then operated on according to Yates method to produce a succession of columns $j = 2, \ldots, \ell$. The successive columns for any k^{th} row are computed as follows:

$$y_{k,j} = y_{i+1,j-1} + y_{i,j-1} \begin{cases} i = 1, 3, 5, \ldots, 2^{\ell} - 1 \\ k = (i + 1)/2 \end{cases}$$

$$y_{k,j} = y_{i+1,j-1} - y_{i,j-1} \begin{cases} i = 1, 3, 5, \ldots, 2^{\ell} - 1 \\ k = (2^{\ell} + i + 1)/2 \end{cases}$$

New columns are computed according to the two preceding equations for $j = 2, \ldots, \ell$ (to create ℓ columns).

Section 3 - Construction and ordering of mean squares. - The mean squares are formed from the parameter estimates (for those terms beyond β_1) and a pointer function is created within the loop DO 309, I = 1, NC.

The array of pointers to the B(I) array is created by the statement IND(I) = I. This array will serve to identify the coefficients in the B(I) array after the process of ordering mean squares according to rank. The ordering is done in the sequence of statements ending with 313.

Operations thus far created a column of mean squares Z(J) with mean squares indexed on J in the order of increasing rank, together with a column of integers IND(J) indexed on J. Thus, any address J will lead to a mean square Z(J) and also to the integer IND(J). This integer is the index I that the associated regression coefficient has in the original Yates order.

The computation of the sums of squares begins following statement 313 and ends with statement 355. The operations are outlined by figure 3(a). The computation of the sums of squares depends on the values of n_0 and m_p according to cases b, c, d, e, and f of table 1(a). Three combinations of these cases are identified in the statement immediately preceding statement 320. If $n_0 = 0$, the situation is that of case a or d and control is transferred to statement 330, and, subsequently, the SS_d of equation (B7) is evaluated at statement 365.

If $n_0 = 1$, the situation can be that of case b ($n_0 = 1$, $m_p = 0$) or case e ($n_0 = 1$, $m_p > 0$) and the SS_b of equation (B4) is computed at the statement following 325.

If $n_0 > 1$, the situation can be that of case c or case f. The quantity SS_c of equation (B5) is computed at the statement for TEM that follows statement 322.

In terms of mathematical symbols previously defined, the strategy parameters are functions of numbers that are read at statement 362 as follows:

Argument FORTRAN symbol	Function Mathematical symbol
MP	m_p
RF	r_F
KPF	α_F
KPU	α_U
RETA	r_η

More than one model-deletion strategy can be used during any computer run. Upon completion of the use of a particular strategy, control is transferred back to statement 362 for the reading of an additional strategy data card. The operation of the program ends when such cards are exhausted.

If m_p and n_0 are each zero, there can be no sequential deletion; therefore, control is transferred to statement 432, and all terms are retained. Setting both α_F and α_U equal to 1.00 (by setting KPF = KPU = 11) is used as a code to signify that no conditional pooling is to be done but that arbitrary deletion is to be accomplished according to values assigned to m_p and r_η. This is done by transferring control to statement 421.

Section 4 - Deletion of terms. - The flow chart for the tests of significance is shown by figure 3(b). The procedure ends at statement 419 (appendix D). The significance tests will have been avoided by earlier statements in section 3 if either $n_0 + m_p = 0$ or both $\alpha_F = 1.0$ and $\alpha_U = 1.0$. Thus, sequential deletion requires both $n_0 + m_p > 0$ and at least one of α_F or $\alpha_U < 1.0$. If $\alpha_F = 1.0$, control is transferred to the U_j test which begins at statement 418. If $\alpha_F < 1.0$, control is determined (fig. 3(b)) by the questions: Is $j > r_F n_0$? Is $j > 1$? If both are yes, control is transferred to statement 418 which initiates the U_j-testing.

Irrespective of whether significance testing is against the F-distribution or the U_j-distribution, insignificance pools $Z_{(j)}$ into the denominator of the test statistic and then transfers control to statement 419 which increases j by one unit. Significance at any j transfers control to statement 420.

The third statement in section 4 (namely, IF (KPF. GT. 10) GO TO 418) transfers control to the U_j-test merely provided $\alpha_F = 1.0$, even if $j < 2$. But the rationale of the U_j-distribution leaves U_j undefined for $j < 2$. The possibility of a transfer of control to the U_j-distribution with $j = 1$ was provided for by setting the critical values of U_j equal to 2.0 for $j = 1$ and all values of $\alpha_U < 1.0$. Thus, if $j = 1$, then obviously $m_p = 0$ and the test statistic is (from table 1(a))

$$u_1 = \frac{2.0z_{(1)}}{SS_b + z_{(1)}} = \frac{2.0}{1 + SS_b/z_{(1)}}$$

for case b ($n_0 = 1$, $m_p = 0$) or

$$u_1 = \frac{n_0 z_{(1)}}{SS_c + z_{(1)}}$$

for case c ($n_0 > 1$, $m_p = 0$). Thus, for case b, $u_1 \leq 2.0$ for $SS_b \geq 0.0$ and $z_{(1)}$ would not test as significant. For case c ($n_0 > 1$, $m_p = 0$), $u_1 > 2.0$ only if

$$\frac{n_0 z_{(1)}}{SS_c + z_{(1)}} > 2.0$$

hence only if

$$z_{(1)} > \frac{2.0SS_c}{(n_0 - 2.0)}$$

Let σ^2 be estimated by $SS_c/(n_0 - 1)$. Then $u_1 > 2.0$ only if

$$z_{(1)} > \frac{2.0(n_0 - 1)}{(n_0 - 2.0)} \frac{SS_c}{(n_0 - 1)} > C(n_0)\hat{\sigma}^2$$

The following table shows $C(n_0)$ as a function of n_0 for the values of n_0 appropriate to case c:

n_0	$C(n_0)$
2	∞
3	4.0
4	3.0
5	2.67
6	2.5

Thus, $z_{(1)}$ (the smallest of the ordered mean squares) would have to be much larger than $\hat{\sigma}^2$ before $z_{(1)}$ would be declared significant.

The flow chart for the model deletion and for the estimate $\hat{\rho}$ is shown by figure 3(c). Transfer of control to statement 420, 421, or completion of the loop at 419 leads to the estimate, respectively,

$$\hat{n} = \text{maximum integer} \leq r_n(j - 1)$$

or

$$\hat{n} = \text{maximum integer} \leq r_n m_p$$

or

$$\hat{n} = \text{maximum integer} \leq r_n(n_c - 1)$$

With $\hat{n}$ so estimated, the $\hat{n}$ smallest absolute value coefficients (beyond b_1^*) are set equal to zero with the statements ending at 425.

APPENDIX D

LISTING OF COMPUTER PROGRAM POOL10

```
C
C     1A.- DECLARATIONS AND TABLES
C
      DIMENSION REMARK(14), ALPHA(11), FTB(20,10), TB(16,10), DATAID(24)
     A, YIN(24,6), IND(16), Z(16), Y (24), B(16), BOUT(16)
C
      DATA (ALPHA(I),I=1,11)/0.001,0.002,0.005,0.01,0.025,0.05,0.10,
     10.25,0.50,0.75,1.0/
C
      DATA((FTB(I,J),J=1,10),I=1,20)/405280.0,101321.3,16211.0,4052.2,64
     A7.8,161.4,39.86,5.828,1.000,0.1716,998.5,498.5,198.5,98.50,38.51,1
     B8.51,8.526,2.571,0.6667,0.1333,167.0,104.3,55.55,34.12,17.44,10.13
     C,5.538,2.024,0.5851,0.1220,74.14,51.45,31.33,21.20,12.22,7.709,4.5
     D45,1.807,0.5486,0.1165,47.18,34.73,22.78,16.26,10.01,6.608,4.060,1
     E.692,0.5281,0.1134,35.51,27.12,18.64,13.74,8.813,5.987,3.776,1.621
     F,0.5149,0.1113,29.24,22.90,16.24,12.25,8.073,5.591,3.589,1.573,0.5
     G057,0.1099,25.42,20.26,14.69,11.26,7.571,5.318,3.458,1.538,0.4990,
     H0.1088,22.86,18.46,13.61,10.56,7.209,5.117,3.360,1.512,0.4938,0.10
     I80,21.04,17.17,12.83,10.04,6.937,4.965,3.285,1.492,0.4897,0.1073,1
     J9.69,16.20,12.23,9.646,6.724,4.844,3.225,1.475,0.4864,0.1068,18.64
     K,15.44,11.75,9.330,6.554,4.747,3.176,1.461,0.4837,0.1063,17.82,14.
     L84,11.37,9.074,6.414,4.667,3.136,1.450,0.4814,0.1059,17.14,14.34,1
     M1.06,8.862,6.298,4.600,3.102,1.4403,0.4794,0.1056,16.59,13.94,10.8
     N0,8.683,6.200,4.543,3.073,1.432,0.4778,0.1053,16.12,13.59,10.58,8.
     O531,6.115,4.494,3.048,1.425,0.4763,0.1051,15.72,13.29,10.38,8.400,
     P6.042,4.4513,3.026,1.419,0.4750,0.1049,15.38,13.03,10.22,8.285,5.9
     Q78,4.414,3.007,1.4130,0.4738,0.1047,15.08,12.81,10.07,8.185,5.922,
     R4.381,2.990,1.408,0.4728,0.1045,14.82,12.62,9.944,8.096,5.872,4.35
     S1,2.975,1.404,0.4719,0.1044/
C
      DATA ((TB(I,J),J=1,10),I=1,16)/2.0,2.0,2.0,2.0,2.0,2.0,2.0,2.0,2.0
     1,2.0,2.0000,1.99999,1.99997,1.99986,1.99917,1.99687,1.9877,1.923,1
     2.706,1.382,2.9976,2.9960,2.9904,2.9809,2.951,2.904,2.806,2.527,2.0
     386,1.688,3.976,3.962,3.925,3.870,3.760,3.625,3.412,2.949,2.395,1.9
     461,4.887,4.845,4.758,4.65,4.44,4.21,3.89,3.287,2.658,2.184,5.74,5.
     563,5.46,5.31,4.99,4.68,4.28,3.57,2.893,2.371,6.51,6.33,6.11,5.87,5
     6.46,5.09,4.61,3.83,3.11,2.54,7.20,6.96,6.65,6.35,5.88,5.44,4.91,4.
     706,3.29,2.69,7.81,7.52,7.10,6.78,6.26,5.75,5.17,4.27,3.45,2.82,8.3
     84,8.01,7.53,7.17,6.59,6.03,5.41,4.45,3.60,2.95,8.82,8.44,7.95,7.53
     9,6.89,6.28,5.61,4.62,3.74,3.07,9.26,8.84,8.33,7.87,7.13,6.50,5.81,
     A4.77,3.87,3.17,9.67,9.21,8.68,8.16,7.37,6.71,5.99,4.92,3.99,3.27,1
     B0.05,9.55,8.95,8.42,7.59,6.91,6.15,5.05,4.10,3.37,10.40,9.86,9.20,
     C8.66,7.79,7.07,6.30,5.17,4.20,3.46,10.72,10.14,9.43,8.83,7.96,7.23
     D,6.44,5.29,4.30,3.55/
```

```
C
C      1B.- INPUTS AND CONSTANTS
C
       READ(5,800) (REMARK(I),I=1,14)
       WRITE(6,801) (REMARK(I),I=1,14)
       READ  (5,802) NO, KODE
       KK = 4
       NC = 16
       WRITE (6,803) NC, NO, KODE
       NCM1 = 15
       NCM2 = 14
       NCP1 = 17
       NT = NC + NO
       FNC = NC
       FNO = NO
       WRITE (6,805)
       DO 103 I=1,NT
       NYIN = 1
       READ (5,804) DATAID(I), NYIN, (YIN(I,K), K=1,NYIN)
       WRITE (6,807) DATAID(I), (YIN(I,K), K=1,NYIN)
       Y(I) = 0.0
       FNYIN = NYIN
       DO 102 K=1,NYIN
       IF (KODE .LT. 1) GO TO 102
       YIN(I,K) = ALOG10(YIN(I,K))
   102 Y(I) = Y(I)+YIN(I,K)
       Y(I) = Y(I)/FNYIN
   103 CONTINUE
       WRITE (6,813)
       DO 105 I=1,NT
       WRITE (6,815) DATAID(I), Y(I)
   105 CONTINUE
C
C      2.- YATES METHOD
C
       II = 2**KK
       IIDB2 = II/2
       KKM1 = KK-1
       DO 208 K=1,KKM1
       DO 206 I=1,II,2
       IP1D2 = (I+1)/2
       B(IP1D2) = Y(I+1)+Y(I)
       LL = IP1D2+IIDB2
   206 B(LL) = Y(I+1)-Y(I)
       DO 207 I=1,II
   207 Y(I) = B(I)
   208 CONTINUE
       DO 209 I=1,II,2
       IP1D2 = (I+1)/2
       B(IP1D2) = Y(I+1)+Y(I)
       LL = IP1D2+IIDB2
       B(LL) = Y(I+1)-Y(I)
   209 CONTINUE
C
```

```
C      3.- CONSTRUCTION AND ORDERING OF MEAN SQUARES
C
       DO 309 I=1,NC
       IND(I)= I
       Z(I) = B(I+1)*B(I+1)/FNC
       B(I)    = B(I)/FNC
   309 CONTINUE
C
       DO 313 J=1,NCM2
       TEST = Z(NCM1)
       IN = NCM1
       DO 312 NA=J,NCM2
       IF(TEST-Z(NA))  312,312,311
   311 TEST = Z(NA)
       IN= NA
   312 CONTINUE
       ITEM= IND(IN)
       TEM= Z(IN)
       IND(IN)= IND(J)
       Z(IN)= Z(J)
       IND(J)= ITEM
       Z(J) = TEM
   313 CONTINUE
C
       TY = 0.0
       IF (NO    -1) 330, 325, 320
   320 SSQYO = 0.0
       DO 322 I=NCP1,NT
       TY = TY + Y (I)
       SSQYO = SSQYO + Y (I)**2
   322 CONTINUE
       TEM1= SSQYO - ((TY**2)/FNO)
       NDF1= NO    -1
       B(1) = (FNC*B(1) + TY)/(FNC + FNO)
       FNDF = NDF1
       S1 = (SQRT(TEM1))/FNDF
       GO TO 355
   325 TY = Y (NCP1)
       TEM1= 0.941176*((TY - B   (1))**2)
       NDF1= 1
       B(1) = (FNC*B(1) + TY)/(FNC + FNO)
       S1 = SQRT(TEM1)
       GO TO 355
   330 TEM1= 0.0
       NDF1= 0
       S1 = 0.0
C
   355 WRITE (6,817) NDF1, S1
C
C         STRATEGY
C
   362 READ (5,808,END = 899) MP,KPF,KPU,RF,RETA
       MPP1 = MP+1
```

```
      NDF  = NDF1
      TEM = TEM1
      RFNO = RF*FNO
      IF (MP .LT. 1) GO TO 370
      DO 365 J = 1,MP
      TEM = TEM + Z(J)
  365 CONTINUE
      NDF = NDF + MP
      FNDF = NDF
      S  = (SQRT(TEM))/FNDF
  370 IF (NO .LT. 1 .AND. MP .LT. 1) GO TO 432
      IF (KPF .GT. 10 .AND. KPU .GT. 10) GO TO 421
C
C     4.- DELETION OF TERMS
C
      DO 419 J=MPP1,NCM1
      JN  = J
      IF (KPF .GT. 10) GO TO 418
      FJ = JN
      IF (FJ .GT. RFNO .AND.   JN  .GT. 1) GO TO 418
C
C           F - TEST
C
      FNDF = NDF
      TEST = FNDF * Z(J) / TEM
      IF (TEST .GT. FTB(NDF,KPF)) GO TO 420
      TEM = TEM + Z(J)
      NDF = NDF + 1
      GO TO 419
C
C          UJ - TEST
C
  418 IF (KPU .GT. 10) GO TO 420
      NDFP1 = NDF + 1
      FNDFP1 = NDFP1
      TEST = FNDFP1*Z(J)/(TEM+Z(J))
      IF (TEST .GT. TB(JN,KPU)) GO TO 420
      TEM = TEM + Z(J)
      NDF = NDF + 1
  419 CONTINUE
      JETA = NCM1
      GO TO 422
  420 JETA = JN-1
      GO TO 422
  421 JETA = MP
  422 ETA = JETA
      JETA = IFIX(RETA*ETA)
      FNDF = NDF
      S  = (SQRT(TEM))/FNDF
      GO TO 434
```

```
  432 JETA = 0
      NDF = 0
      S = 0.0
  434 JA = JETA + 1
      IRHO = NCM1 - JETA
      DO 424 I=1,NC
      BOUT(I) = B(I)
  424 CONTINUE
      IF (JETA .LT. 1) GO TO 505
      DO 425 J=1,JETA
      INDX = IND(J) + 1
      BOUT(INDX) = 0.0
  425 CONTINUE
C
C     5.- OUTPUT
C
  505 WRITE (6,809) (MP, RF, ALPHA(KPF), ALPHA(KPU), RETA)
      WRITE (6,811) IRHO, JETA, Z(JA)
      WRITE (6,821)
      DO 510 I=1,NC
      INDXI = I
      WRITE (6,819) INDXI, BOUT(I)
  510 CONTINUE
      WRITE (6,817) NDF, S
      GO TO 362
  899 STOP
C
  800 FORMAT (13A6,A2)
  801 FORMAT (1H1,//10X,13A6,A2//)
  802 FORMAT (2I1)
  803 FORMAT (1H0,3X,4HNC =I5,5X,4HNO =I5,5X,6HKODE =I2)
  804 FORMAT (A6,I6,6F6.0)
  805 FORMAT (1H0,6HDATAID,5X,4HY(1),9X,4HY(2),9X4HY(3),9X,4HY(4),9X,4HY
     A(5),9X,4HY(6))
  807 FORMAT (1X,A6,6E13.5)
  808 FORMAT (3I5,2F5.3)
  809 FORMAT (1H1//,1X,4HMP =I5,5X,4HRF =F6.3,5X,8HALPHAF =F6.3,5X,8HALP
     AHAU =F6.3,5X,6HRETA =F6.3)
  811 FORMAT (1H0,6X,6HIRHO =,I3,9X,6HJETA =,I3,9X,11HZ(JETA+1) =,E13.5)
  813 FORMAT (1H0,6HDATAID,6X,4HY(I))
  815 FORMAT (1X,A6,E13.5)
  817 FORMAT (1H0,7X,5HNDF =,I3,13X,3HS =,E13.5)
  819 FORMAT (1X,I3,E15.5)
  821 FORMAT (1H0,6HDATAID,6X,4HB(I))
      END
```

APPENDIX E

ILLUSTRATIVE OUTPUT OF COMPUTER PROGRAM POOL10

SECOND FACTORIAL EXPERIMENT, FULL REPLICATE OF 2**4   04-02-82

NC =   16   NO =   4   KODE = 1

DATAID	Y(1)	Y(2)	Y(3)	Y(4)	Y(5)	Y(
1	.12670+03	.17650+03				
2	.19600+03	.18410+03				
3	.16340+03	.15260+03				
4	.19400+03	.24940+03				
5	.88900+02	.10610+03				
6	.17580+03	.16010+03				
7	.15490+03	.18220+03				
8	.14410+03	.16240+03				
9	.13610+03	.10700+03				
10	.65700+02	.60000+02				
11	.12980+03	.10720+03				
12	.80600+02	.87700+02				
13	.17580+03	.16480+03				
14	.16720+03	.16610+03				
15	.14180+03	.12920+03				
16	.14500+03	.14010+03				
C-1	.19520+03	.18560+03				
C-2	.11430+03	.88000+02				
C-3	.14950+03	.11980+03				
C-4	.15870+03	.14320+03				

DATAID	Y(I)
1	.21748+01
2	.22787+01
3	.21984+01
4	.23423+01
5	.19873+01
6	.22247+01
7	.22253+01
8	.21846+01
9	.20816+01
10	.17979+01
11	.20717+01
12	.19247+01
13	.22310+01
14	.22218+01

5	.21315+01
)	.21539+01
-1	.22795+01
-2	.20013+01
-3	.21265+01
-4	.21783+01

NDF = 3 S = .66830-01

= 0 RF = .500 ALPHAF = .500 ALPHAU = .500 RETA = .800

IRHO = 6 JETA = 9 Z(JETA+1) = .11194-01

TAID	B(I)
1	.21408+01
2	.00000
3	.00000
4	.00000
5	.30628-01
6	.00000
7	.00000
8	-.26450-01
9	-.62629-01
0	-.53884-01
1	.00000
2	.00000
3	.77157-01
4	.30950-01
5	.00000
6	.00000

NDF = 15 S = .23759-01

APPENDIX F

SYMBOLS

Mathematical symbol	FORTRAN name	Description
b_i	B(I)	estimate of β_i
$C_{ae,mx}$		adjusted coefficient of error
$E(. . .)$		expectation of . . .
e		single observation random error
$\bar{e}^2_{max}$		maximum over hypercube of mean square prediction error over simulations
g		number of independent variables
$g-h$		experiment contains 2^{g-h} treatments
h		experiment contains $(1/2)^h$ times number of treatments in full-factorial experiment
i	I	subscript denoting order of computing mean squares according to Yates' algorithm: $i = 1, 2, \ldots, n_c$, I = 1, ..., NC
j	J	subscript denoting the j smallest mean square (exclusive of grand mean): $j = 1, 2, \ldots, n_c-1$
k	K	subscript
	KPF	index number for α_F
	KPU	index number for α_U
m_p	MP	number of mean squares pooled before testing begins
	NYIN	number of observations averaged for each treatment (maximum of six)
n_0	NO	number of center points
n_c	NC	number of hypercube points
n_t	NT	total number of observations in one experiment
r_F	RF	distribution transfer parameter

r_η	RETA	number of terms deleted is integer value of r_η times number insignificant
U_j	TB	test statistic
$V(. . .)$		variance of . . .
x_k		k^{th} independent variable
Y		conceptual value of dependent variable
$\hat{Y}$		estimate of response function from fitted model
y_i	YIN(I,K)	response variate: I = 1, ..., NT K = 1, ..., NYIN
Z_i	Z(I)	mean squares in Yates' order
α_F		nominal significance level of F test
α_U		nominal significance level of U_j test
β_i		coefficients of model equation that are estimated in Yates order
η		number of mean squares having noncentrality parameter of zero
$\hat{\eta}$	JETA	number of terms deleted
ν		coefficient of variation
ρ		number of real effects (not including zero degree coefficient)
$\hat{\rho}$	IRHO	number of terms retained (beyond x^0)
σ		standard deviation of e

REFERENCES

Birnbaum, A. (1959). On the Analysis of Factorial Experiments Without Replication. Technometrics 1: 343-357.

Box, G. E. P., and Wilson, K. B. (1951). On the Experimental Attainment of Optimum Conditions. J. Roy. Stat. Soc., Ser. B, 13: 1-45.

Box, G. E. P., and Hunter, J. S. (1957). Multi-Factor Experiment Designs for Exploring Response Surfaces. Ann. Math. Stat., 28: 195-241.

Cochran, W. G. (1941). The Distribution of the Largest of a Set of Estimated Variances as a Fraction of Their Total. Ann. Eugenics, 11: 47-52.

Collins, H. E., Quigg, R. J., and Dreshfield, R. L. (1968). Development of a Nickel-Base Superalloy Using Statistically Designed Experiments. ASM Trans., 61: 771-721.

Daniel, C. (1959). Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments. Technometrics 1: 311-341.

Davies, Owen L., ed. (1956). Design and Analysis of Industrial Experiments. Second ed., Hafner Publishing Co.

Filippi, A. M. (1974). TD Ni Cr (Ni-20 Cr-2ThO_2) Forging Studies. (WANL-M-FR-74-001, Westinghouse Astro-nuclear Lab.; NASA Contract NAS3-15548) NASA CR-134619.

Holms, Arthur G. (1967). Designs of Experiments as Telescoping Sequences of Blocks for Optimum Seeking (as intended for alloy development). NASA TN D-4100.

Holms, A. G., and Berrettoni, J. N. (1967). Multiple Decision Procedures for ANOVA of Two-Level Factorial Fixed-Effects Replication-Free Experiments. NASA TN D-4272.

Holms, A. G., and Berrettoni, J. N. (1969). Chain Pooling ANOVA for Two-Level Factorial Replication-Free Experiments. Technometrics, 11: 725-746.

Holms, Arthur G. (1974). Chain Pooling to Minimize Prediction Error in Subset Regression. NASA TM X-71645.

Holms, Arthur G. (1977). "Chain Pooling" Model Selection as Developed for the Statistical Analysis of a Rotor Burst Protection Experiment. NASA TM-73874.

Holms, Arthur G. (1979). Numbers of Center Points Appropriate to Blocked Response Surface Experiments. NASA TM-79201.

Holms, Arthur G. (1980). Backward Deletion to Minimize Prediction Errors in Models from Factorial Experiments with Zero to Six Center Points. NASA TM-81524.

Kennedy, W. J., and Bancroft, T. A. (1971). Model Building for Prediction in Regression Based Upon Repeated Significance Tests. Ann. Math. Stat., 42: 1273-1284.

Kent, William B. (1972). Development Study of Compositions for Advanced Wrought Nickel-Base Super Alloys. NASA CR-120934.

Lucas, James M. (1974). Optimum Composite Designs. Technometrics, 16: 561-567.

Lucas, James M. (1976). Which Response Surface Design is Best Technometrics 18: 411-417.

Sandrock, Gary D., and Holms, Arthur G. (1969). Statistical Design and Analysis of Optimum Seeking Experiments to Develop a Gamma-Prime Strengthened Cobalt-Nickel Base Alloy. NASA TN D-5587.

Walls, R. C., and Weeks, D. L. (1969). A Note on the Variance of a Predicted Response in Regression. Am. Stat., 23: 23-26.

Wilk, M. B., Gnanadesikan, R., and Freeny, Anne E. (1963). Estimation of Error Variance from Smallest Ordered Contrasts. J. Am. Stat. Assoc., 58:/152-160.

Zahn, Douglas A. (1975). An Empirical Study of the Half-Normal Plot. Technometrics 17: 201-211.

TABLE 1. - CHAIN-POOLING DELETION

(a) Expressions for degrees of freedom, sums of squares, and test statistics

Case	n_0	m_p	n_{df0}	SS_0	Eq. no.	F-Test (see eq. (14)) $(1 \le j \le r_f n_0)$	U -Test (see eq. (16)) $(1 \leqq j \leqq n_c - 1)$
a	0	0					
b	1	0	1	$SS_b = 0.941176(y_{01} - b_1)^2$	(B4)	$f_{i,j} = \dfrac{(1 + j - 1)z_{(j)}}{SS_b + z(1) + \ldots + z(j-1)}$	$u_j = \dfrac{(n_{df0} + j)z_{(j)}}{SS_b + z_{(1)} + \ldots + z_{(j)}}$
c	>1	0	$n_0 - 1$	$SS_c = \sum_{k=1}^{n_0} y_{0k}^2 - \frac{1}{n_0}\left(\sum_{k=1}^{n_0} y_{0k}\right)^2$	(B5)	$f_{1,n_0+j-2} = \dfrac{(n_0 - 1 + j - 1)z_{(j)}}{SS_c + z_{(1)} + \ldots + z_{(j-1)}}$	$u_j = \dfrac{(n_{df0} + j)z_{(j)}}{SS_c + z_{(1)} + \ldots + z_{(j)}}$
d	0	>0	m_p	$SS_d = \sum_{j=1}^{m_p} z(j)$	(B7)		$u_j = \dfrac{jz_{(j)}}{SS_d + z_{(mp+1)} + \ldots + z_{(j)}}$
e	1	>0	$1 + m_p$	$SS_e = SS_b + SS_d$	(B9)	$f_{1,j} = \dfrac{(1 + j - 1)z_{(j)}}{SS_e + z_{(mp+1)} + \ldots + z_{(j-1)}}$	$u_j = \dfrac{(1 + j)z_{(j)}}{SS_e + z_{(mp+1)} + \ldots + z_{(j)}}$
f	>1	>0	$n_0 - 1 + m_p$	$SS_f = SS_c + SS_d$	(B11)	$f_{1,n_0+j-2} = \dfrac{(n_0 - 1 + j - 1)z_{(j)}}{SS_f + z_{(mp+1)} + \ldots + z_{(j-1)}}$	$u_j = \dfrac{(n_0 - 1 + j)z_{(j)}}{SS_f + z_{(mp+1)} + \ldots + z_{(j)}}$

TABLE 1. - Continued.

(b) Values of j at which transfer from F- to U_j-distribution occurs

r_F	n_0						
	0	1	2	3	4	5	6
	j						
0.0	2	2	2	2	2	2	2
.4	2	2	2	2	2	3	3
.5	2	2	2	2	3	3	4
.7	2	2	2	3	3	4	5
.9	2	2	2	3	4	5	6
1.0	2	2	3	4	5	6	7
2.0	2	3	5	7	9	11	13
4.0	2	5	9	13	n.t.[a]	n.t.	n.t.
8.0	2	9	n.t.	n.t.	n.t.	n.t.	n.t.
16.0	2	n.t.	n.t.	n.t.	n.t.	n.t.	n.t.

[a]No transfer occurs, n.t.

TABLE 1. - Continued.

(c) Some admissible strategies and their operating characteristics

n_0	Strategy	m_p	r_F	α_F	α_U	r_n	$C_{ae,mx}(\nu)$		$R(\nu)$	
							ν = 64.3	ν = 4.0	ν = 64.3	ν = 4.0
0	a	0	---	1.0	1.0	0.0	33.04	2.065	1.1215	1.0000
	b	1	---	1.0	.50	.25	32.95	2.240	1.1185	1.0847
	c	5	---	1.0	.05	.75	29.46	24.09	1.0000	11.6659
1	a	0	2.0	.50	.75	.65	33.05	2.118	1.0875	1.0000
	b	0	3.0	.50	.10	.80	31.78	2.180	1.0457	1.0293
	c	0	8.0	.005	.05	.775	30.39	24.50	1.0000	11.5675
2	a	0	2.0	.75	.50	.90	34.40	2.187	1.1058	1.0000
	b	0	3.0	.25	.50	.85	33.14	2.252	1.0653	1.0297
	c	0	4.0	.002	.05	.80	31.11	25.41	1.0000	11.6187
3	a	0	1.0	.75	.50	.85	34.41	2.244	1.0770	1.0000
	b	0	.0	.75	.50	.80	33.29	2.301	1.0419	1.0254
	c	0	2.0	.001	.025	.75	31.95	26.12	1.0000	11.6399
4	a	0	.5	.75	.50	.85	34.38	2.300	1.0536	1.0000
	b	0	.5	.50	.50	.80	33.89	2.309	1.0386	1.0039
	c	0	.9	.10	.05	.80	32.63	15.11	1.0000	6.5696
5	a	0	.6	.50	.10	.60	35.92	2.352	1.0719	1.0000
	b	0	1.0	.25	.10	.80	33.97	2.394	1.0137	1.0179
	c	0	.5	.25	.05	.80	33.51	9.744	1.0000	4.1429
6	a	0	.6	.50	.10	.60	36.77	2.401	1.0736	1.0000
	b	0	.5	.50	.05	.80	34.72	2.408	1.0137	1.0029
	c	0	.0	.50	.025	.80	34.25	8.384	1.0000	3.4919

[a]Minimizes $C_{ae,mx}(4.0)$.
[b]Security regret.
[c]Minimizes $C_{ae,mx}(64.3)$.

TABLE 1. - Continued.

(d) Results for F-only strategies

n_0	α_F	$\bar{\rho}$		$C_{ae,mx}(\nu)$		$R(\nu)$	
		ν = 64.3	ν = 4.0	ν = 64.3	ν = 4.0	ν = 64.3	ν = 4.0
1	0.05	3.760	7.627	38.09	21.60	------	------
	.10	6.576	11.58	32.99	6.403	1.0000	3.0189
	.25	9.858	13.68	33.20	2.263	1.0064	1.0669
	.50	11.87	14.40	33.63	2.166	1.0194	1.0212
	.75	13.16	14.72	33.89	2.122	1.0273	1.0005
	1.00	15.00	15.00	33.94	2.121	1.0288	1.0000
2	0.05	3.462	7.497	39.51	22.56	------	------
	.10	6.398	11.58	33.81	6.387	1.0000	2.9191
	.25	9.737	13.63	34.34	2.331	1.0157	1.0654
	.50	11.83	14.38	34.70	2.219	1.0263	1.0142
	.75	13.14	14.70	34.95	2.191	1.0337	1.0014
	1.00	15.00	15.00	35.00	2.188	1.0352	1.0000
3	0.05	2.677	9.294	41.07	15.24	------	------
	.10	5.104	12.23	35.46	4.494	------	------
	.25	8.471	13.74	34.87	2.353	1.0000	1.0486
	.50	10.95	14.36	35.46	2.283	1.0169	1.0174
	.75	12.72	14.70	35.83	2.247	1.0275	1.0013
	1.00	15.00	15.00	35.91	2.244	1.0298	1.0000
4	0.05	2.235	10.33	42.01	10.96	------	------
	.10	4.360	12.48	36.68	3.799	------	------
	.25	7.685	13.74	35.37	2.415	1.0000	1.0495
	.50	10.44	14.36	36.22	2.335	1.0240	1.0148
	.75	12.46	14.70	36.66	2.305	1.0365	1.0017
	1.00	15.00	15.00	36.81	2.301	1.0407	1.0000
5	0.05	2.020	11.03	43.62	8.328	------	------
	.10	3.931	12.71	37.64	3.429	------	------
	.25	7.232	13.74	36.03	2.476	1.0000	1.0527
	.50	10.09	14.36	36.92	2.385	1.0247	1.0140
	.75	12.34	14.71	37.48	2.356	1.0402	1.0017
	1.00	15.00	15.00	37.64	2.352	1.0447	1.0000
6	0.05	1.881	11.52	44.76	6.195	------	------
	.10	3.653	12.82	39.00	3.139	------	------
	.25	6.885	13.75	36.63	2.515	1.0000	1.0466
	.50	9.871	14.36	37.77	2.431	1.0311	1.0117
	.75	12.21	14.69	38.43	2.407	1.0491	1.0017
	1.00	15.00	15.00	38.46	2.403	1.0500	1.0000

TABLE 1. - Concluded.

(e) Comparison of F-only results for $C_{ae,mx}(\nu)$ with multiparameter results

	n_0											
	1		2		3		4		5		6	
	ν											
	64.3	4.0	64.3	4.0	64.3	4.0	64.3	4.0	64.3	4.0	64.3	4.0
a	31.78	2.180	33.14	2.252	33.29	2.301	33.89	2.309	33.97	2.394	34.72	2.408
b	38.09	21.60	39.51	22.56	41.07	15.24	42.01	10.96	43.62	8.328	44.76	6.195
c	19.8	890.8	19.2	901.8	23.4	562.3	24.0	374.7	28.4	247.9	28.9	157.3
d	33.63	2.166	34.70	2.219	35.46	2.283	36.22	2.335	36.92	2.385	37.77	2.431
e	5.8	-.6	4.7	-1.5	6.5	-.8	6.9	1.1	8.7	-.4	8.8	1.0

[a] $C_{ae,mx}$ for multiparameter strategy.
[b] $C_{ae,mx}$ for F-only strategy, $\alpha = 0.05$.
[c] Percent increase of F-only (0.05) over multiparameter.
[d] $C_{ae,mx}$ for F-only strategy, $\alpha = 0.50$.
[e] Percent increase of F-only (0.50) over multiparameter.

TABLE 2. - FIRST FACTORIAL EXPERIMENT, ONE-HALF

REPLICATE OF 2^5

[Contents of Ni, W, and Zr held constant at 38, 14, and 0.25 wt. %, respectively; balance, Co.]

(a) Levels of independent variables in natural units

Variable	Design center	Lower level	Upper level	Scale factor
	Level, wt.%			
ξ_{Ti}	1.0	0.5	1.5	0.5
ξ_{Cr}	4.0	2.0	6.0	2.0
ξ_{C}	.4	.3	.5	.1
ξ_{Al}	7.0	6.75	7.25	.25
ξ_{T}	---	---	---	---
°C	1593	1566	1621	28
(°F)	(2900)	(2850)	(2950)	(50)

(b) Levels of independent variables in design units and corresponding results

Alloy	Level					Stress-rupture[a] life, hr	Elongation[a], %
	x_{Ti}	x_{Cr}	x_{C}	x_{Al}	x_{T}		
1	-1	-1	-1	-1	-1	175.1, 199.4	31, 15
2	+1	-1	-1	-1	+1	83.2, 166.5	4, 8
3	-1	+1	-1	-1	+1	22.9, 24.5	9, 8
4	+1	+1	-1	-1	-1	14.7, 21.1	3, 8
5	-1	-1	+1	-1	+1	153.5, 237.6	13, 22
6	+1	-1	+1	-1	-1	119.5, 129.6	10, 10
7	-1	+1	+1	-1	-1	28.2, 39.0	17, 14
8	+1	+1	+1	-1	+1	30.0, 38.1	14, 11
9	-1	-1	-1	+1	+1	55.1, 79.2	11, 10
10	+1	-1	-1	+1	-1	29.2, 47.0	5, 8
11	-1	+1	-1	+1	-1	3.5, 11.1	10, 15
12	+1	+1	-1	+1	+1	17.7, 19.6	14, 13
13	-1	-1	+1	+1	-1	132.1, 190.7	19, 16
14	+1	-1	+1	+1	+1	94.1, 95.1	10, 9
15	-1	+1	+1	+1	+1	12.7, 19.1	21, 10
16	+1	+1	+1	+1	-1	16.7, 16.8	18, 15

[a]At stress of 103 MN/m^2 (15 ksi) and temperature of 1010° C (1850° F).

TABLE 3. - MODEL COEFFICIENT ESTIMATES FROM FIRST FACTORIAL EXPERIMENT

Term	Minimum $C_{ae,mx}(4.0)$	Security regret	Minimum $C_{ae,mx}(64.3)$
	Model coefficient estimates		
x_0	1.652	1.652	1.652
x_{Ti}	-.030	-.030	.000
x_{Cr}	-.383	-.383	-.383
$x_{Ti}x_{Cr}$	.078	.078	.078
x_C	.100	.100	.100
$x_{Ti}x_C$	-.017	-.017	.000
$x_{Cr}x_C$	-.003	-.003	.000
$x_{Al}x_T$	.022	.022	.000
x_{Al}	-.146	-.146	-.146
$x_{Ti}x_{Al}$	.034	.034	.000
$x_{Cr}x_{Al}$	-.002	-.002	.000
x_Cx_T	-.045	-.045	.000
x_Cx_{Al}	.042	.042	.000
$x_{Cr}x_T$	.036	.036	.000
$x_{Ti}x_T$	.052	.052	.052
x_T	.037	.037	.000
ndf	0	2	14
s	.000	.007	.064

TABLE 4. - SECOND FACTORIAL EXPERIMENT, FULL

REPLICATE OF 2^4

[Contents of Ni, W, and Zr held constant at 38, 14, and 0.25 wt.%, respectively; balance, Co.]

(a) Levels of independent variables in natural units

Variable	Design center	Lower level	Upper level	Scale factor
	Level, wt.%			
ξ_{Ti}	1.0	0.5	1.5	0.5
ξ_{Cr}	2.0	1.5	2.5	.5
ξ_{C}	.5	.4	.6	.1
ξ_{Al}	6.75	6.5	7.0	.25

(b) Levels of independent variables in design units and corresponding results

Alloy	Level				Stress-rupture[a] life, hr	Elongation[a], %
	x_{Ti}	x_{Cr}	x_{C}	x_{Al}		
1	-1	-1	-1	-1	126.7, 176.5	13, 12
2	+1	-1	-1	-1	196.0, 184.1	10, 8
3	-1	+1	-1	-1	163.4, 152.6	24, 12
4	+1	+1	-1	-1	194.0, 249.4	6, 10
5	-1	-1	+1	-1	88.9, 106.1	15, 25
6	+1	-1	+1	-1	175.8, 160.1	13, 15
7	-1	+1	+1	-1	154.9, 182.2	17, 14
8	+1	+1	+1	-1	144.1, 162.4	14, 13
9	-1	-1	-1	+1	136.1, 107.0	17, 14
10	+1	-1	-1	+1	65.7, 60.0	10, 8
11	-1	+1	-1	+1	129.8, 107.2	11, 15
12	+1	+1	-1	+1	80.6, 87.7	9, 11
13	-1	-1	+1	+1	175.8, 164.8	19, 15
14	+1	-1	+1	+1	167.2, 166.1	13, 12
15	-1	+1	+1	+1	141.8, 129.2	14, 19
16	+1	+1	+1	+1	145.0, 140.1	13, 12

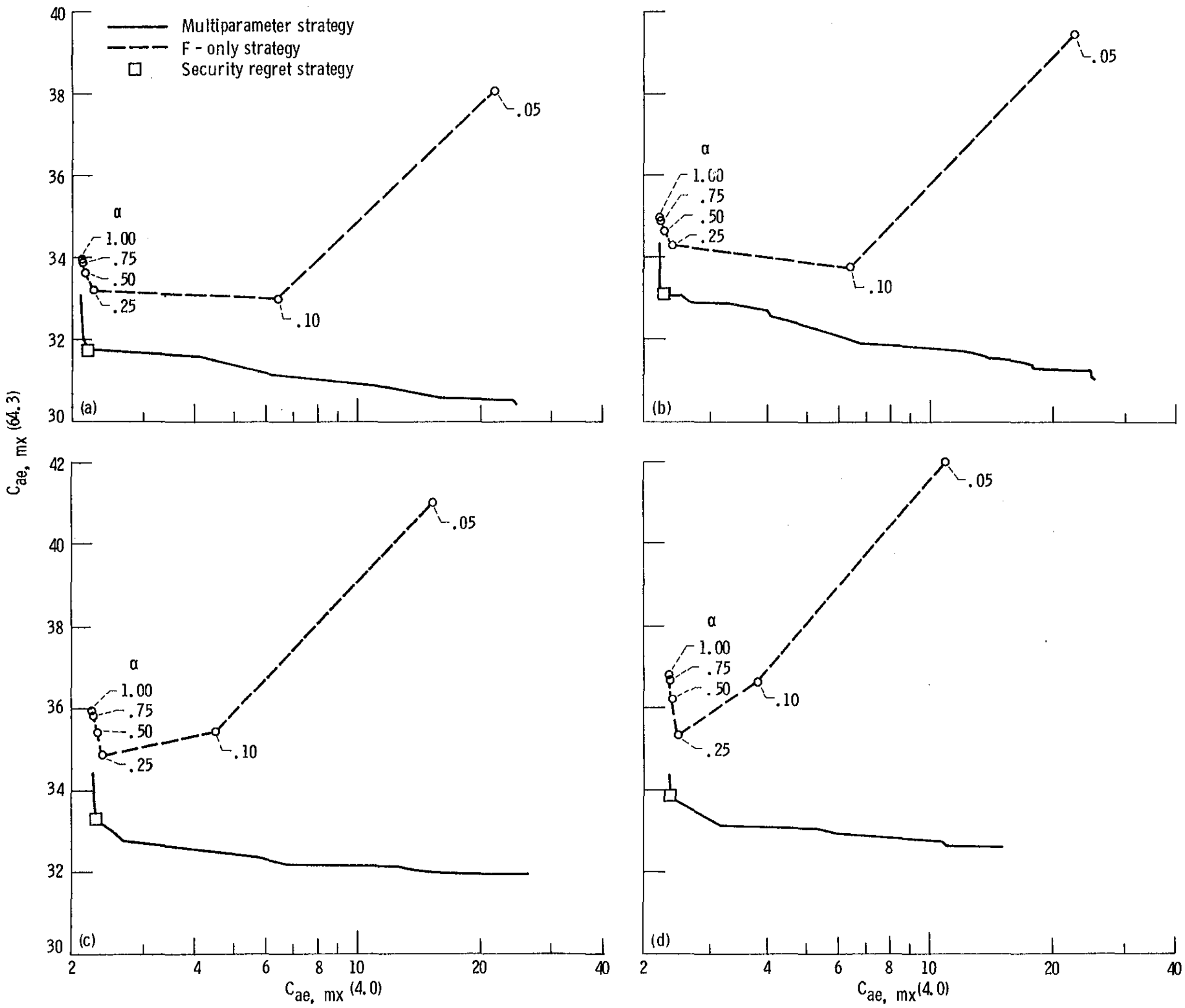
[a]At stress of 103 MN/m^2 (15 ksi) and temperature of 1010° C (1850° F).

TABLE 4. - Concluded.

(c) Results from three deletion strategies

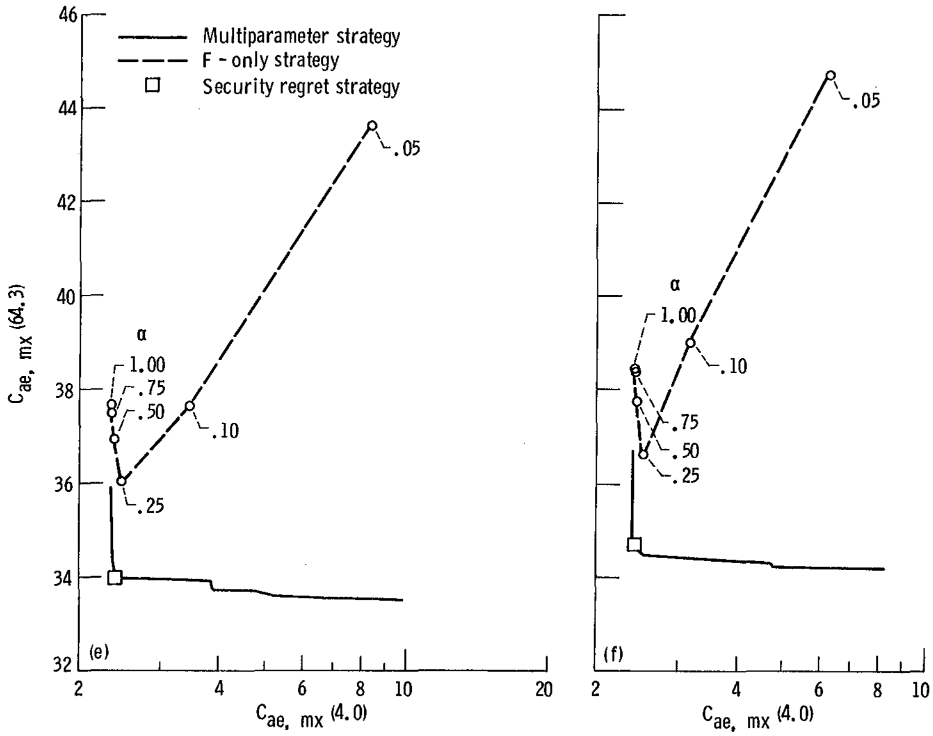
n_0	Minimum $C_{ae,mx}(4.0)$			Security regret			Minimum $C_{ae,mx}(64.3)$		
	$\hat{\rho}$	n_{df}	s	$\hat{\rho}$	n_{df}	s	$\hat{\rho}$	n_{df}	s
0	15	0	0.0	15	1	0.007	4	15	0.036
1	8	13	.025	4	16	.035	4	16	.035
2	14	3	.066	4	15	.032	4	16	.036
3	14	4	.049	6	14	.025	4	17	.034
4	5	15	.024	[a]6	[a]15	[a].024	4	18	.032

[a]See illustrative output, appendix E.



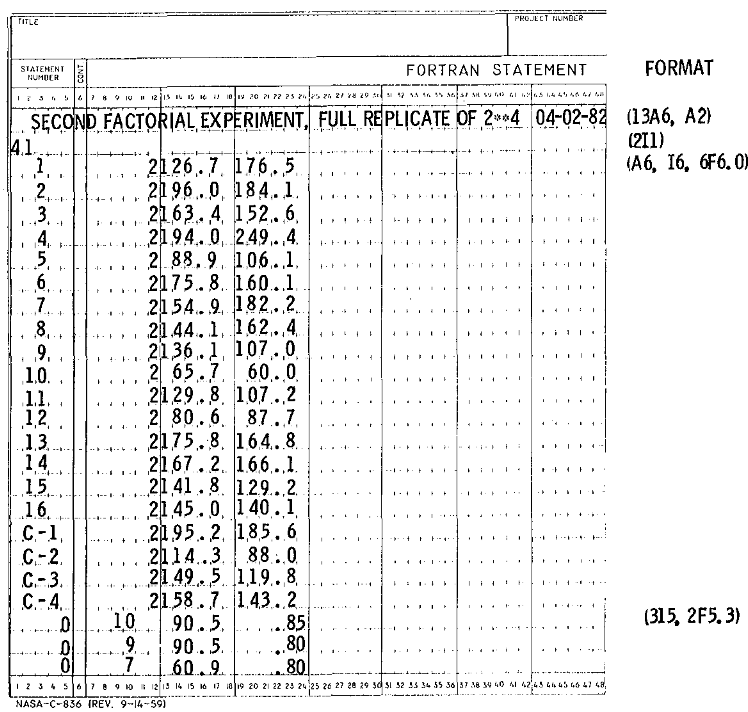
(a) $n_0 = 1$. (b) $n_0 = 2$. (c) $n_0 = 3$. (d) $n_0 = 4$.

Figure 1. - Strategy comparisons.



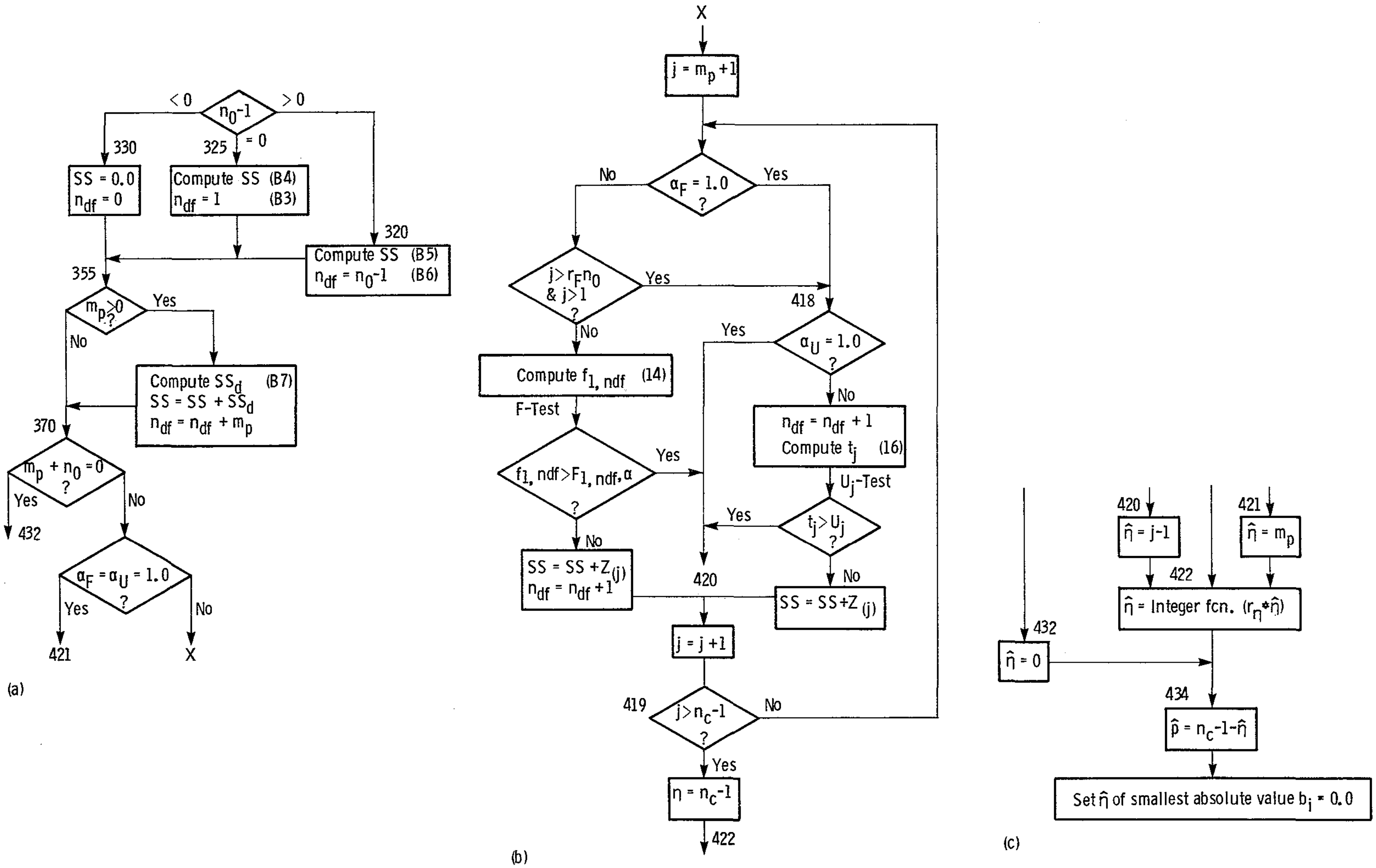
(e) $n_0 = 5$.　　(f) $n_0 = 6$.

Figure 1. - Concluded.



```
TITLE                                                         PROJECT NUMBER
STATEMENT NUMBER | CONT | FORTRAN STATEMENT                                       FORMAT
 SECOND FACTORIAL EXPERIMENT, FULL REPLICATE OF 2**4  04-02-82                    (13A6, A2)
41                                                                                (2I1)
    1           2126.7 176.5                                                      (A6, I6, 6F6.0)
    2           2196.0 184.1
    3           2163.4 152.6
    4           2194.0 249.4
    5           2 88.9 106.1
    6           2175.8 160.1
    7           2154.9 182.2
    8           2144.1 162.4
    9           2136.1 107.0
   10           2 65.7  60.0
   11           2129.8 107.2
   12           2 80.6  87.7
   13           2175.8 164.8
   14           2167.2 166.1
   15           2141.8 129.2
   16           2145.0 140.1
  C-1           2195.2 185.6
  C-2           2114.3  88.0
  C-3           2149.5 119.8
  C-4           2158.7 143.2
        0   10    90.5   .85                                                      (3I5, 2F5.3)
        0    9    90.5   .80
        0    7    60.9   .80
NASA-C-836 (REV. 9-14-59)
```

Figure 2. - Illustrative input, computer program pool 10.



(a) Computation of sums of squares.

(b) Tests of significance.

(c) Deletion of insignificant coefficients.

Figure 3. - Flow chart of program POOLIO. Numbers in ( )'s are equation numbers of text. Three-digit integers are statement numbers in appendix D.

1. Report No. NASA TM-82995	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle COMPUTER PROGRAM TO MINIMIZE PREDICTION ERROR IN MODELS FROM EXPERIMENTS WITH 16 HYPERCUBE POINTS AND ZERO TO SIX CENTER POINTS		5. Report Date November 1982
		6. Performing Organization Code 505-33-12
7. Author(s) Arthur G. Holms		8. Performing Organization Report No. E-1428
		10. Work Unit No.
9. Performing Organization Name and Address National Aeronautics and Space Administration Lewis Research Center Cleveland, Ohio 44135		11. Contract or Grant No.
		13. Type of Report and Period Covered Technical Memorandum
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

A previous report described a backward deletion procedure of model selection that was optimized for minimum prediction error and which used a multiparameter combination of the F-distribution and an order statistics distribution of Cochran's. A computer program is now described that applies the previously optimized procedure to real data. The use of the program is illustrated by examples.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Statistics and probability; Statistical computing; Regression analysis; Subset regression; Factorial experiments; Response surfaces; Chain pooling	Unclassified - unlimited STAR Category 65

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified		

* For sale by the National Technical Information Service, Springfield, Virginia 22161

End of Document